

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



TRABAJO FIN DE MÁSTER

**"Análisis bioinformático del efecto de la
inhibición genética de Ambra1 en fibroblastos
embrionarios transformados"**

**Máster Universitario en Bioinformática y Biología
Computacional**

Autor: Mendiburu-Elicabe Garganta, Marina

Tutor: Velasco Díez, Guillermo
**Departamento: Universidad Complutense/ Instituto de
Investigaciones Sanitarias San Carlos (UCM/IdISSC)**

FECHA: Febrero, 2019

a Ignacio, Daniel y Malena

ÍNDICE

RESUMEN

ABREVIATURAS

OBJETIVOS

1- INTRODUCCIÓN	1
1.1- Autofagia	1
1.2- AMBRA1	3
1.3- Fibroblastos	5
1.4- Transición Epitelio Mesénquima (EMT)	6
1.5-GENE ONTOLOGY (GO)	7
1.6-Kyoto Encyclopedia of Genes and Genomes (KEGG)	9
1.7-REACTOME	10
1.8-Gene Set Enrichment Analysis (GSEA)	11
 2- MATERIALES Y MÉTODOS	 13
2.1- R	13
2.2- Bioconductor	14
2.2.1- Limma	15
2.2.2- gplots	15
2.2.3- ggplot2	15
2.2.4- illuminaMousev2.db	16
2.2.5- goProfiles	16
2.2.6- clusterProfiler	17
2.2.7- ReactomePA	17
2.2.8- DOSE	17
2.2.9- topGO	18
2.2.10- enrichplot	18
2.2.11- GO.db	18
2.2.12- KEGG.db	19
2.2.13- KEGGREST	19
2.2.14- reactome.db	19
 3- RESULTADOS	 20
3.1- Resultados experimentales previos	20
3.2- Análisis del efecto de la inhibición genética de Ambra1	23

3.2.1- GO análisis de los genes expresados diferencialmente	23
3.2.1.1- goProfiles	23
3.2.1.2- clusterProfiler	30
3.2.1.2.1- groupGO	30
3.2.1.2.2- enrichGO	34
3.2.1.2.3- gseGO	52
3.2.2- KEGG análisis de los DEGs	53
3.2.2.1- enrichKEGG	53
3.2.2.2- gseKEGG	57
3.2.3- Reactome pathway database análisis de los DEGs.	57
3.2.3.1- ReactomePA	57
4- DISCUSIÓN	60
5- CONCLUSIONES	69
6-BIBLIOGRAFIA	70
ANEXOS	

RESUMEN

Trabajos previos del grupo donde se ha desarrollado el presente TFM han mostrado la importancia de Ambra1 en cáncer. El objetivo principal de este estudio ha sido el análisis bioinformático del efecto de la inhibición genética de Ambra1 en fibroblastos embrionarios transformados. Para lograr este propósito, hemos realizado un código en R para examinar los datos de un array obtenido del análisis de fibroblastos “wild type” y fibroblastos transformados deficientes en Ambra1 (“KO”).

Específicamente realizamos un análisis mediante Gene Ontology (GO), para desarrollar una representación computacional de como los genes codifican funciones biológicas a nivel molecular, celular y del sistema tisular; mediante KEGG para comprender las funciones de alto nivel y las utilidades del sistema biológico.

Las funciones utilizadas fueron ***basicProfile***, ***groupGO***, ***enrichGO***, ***enrichKEGG***, ***enrichPathway*** obteniendo como resultado la confirmación de la hipótesis de partida: la existencia de una asociación entre la inactivación de Ambra1 y un aumento de la adhesión molecular, pérdida de la polaridad apical-basal en el proceso EMT, apoptosis, señalización a través de la vía de fosfatidilinositol. Además, el análisis desarrollado ha permitido identificar la existencia de una correlación entre la inhibición de Ambra1 y la activación de histonas que implica diversos procesos biológicos como ensamblaje de cromatina, ensamblaje de nucleosoma, empaquetamiento de DNA, ensamblaje/desensamblaje de cromatina, recombinación meiótica, expresión de rRNA que no ha sido descrito con anterioridad.

ABREVIATURAS

Ambra1- molécula de activación BECN1- reguladora autofagia 1

ATG5- autofagia relacionado 5

ATG7- autofagia relacionado 7

BP - proceso biológico

CC - componente o localización sub-celular

CMA- autofagia mediada por chaperonas

CNV- variación de número de copia

DEGs - genes expresados diferencialmente

DOSE - Disease Ontology, ontología de la enfermedad

EMT- Transición Epitelio Mesénquima

FDR- tasa de falsos positivos

GO - Gene Ontology

GSEA - enriquecimiento de conjuntos de genes

KEGG - Enciclopedia de genes y genomas de Kyoto

KO- knock out

LIMMA- Linear Models for Microarray Data

MET- transición mesénquima-epitelio

MF - función molecular

MMPs- metaloproteinasas de matriz

mTOR- diana de Rapamicina en mamíferos

PE- fosfatidiletanolamina

SCC- Carcinoma escamoso de piel

SQSTM1- sequestosome 1, p62

TGFβ1- Factor de Crecimiento Transformante beta 1

TGFβ2- Factor de Crecimiento Transformante beta 2

TGFβ3- Factor de Crecimiento Transformante beta 3

ULK1- unc-51 quinase activadora de autofagia 1

Wt- wild type

OBJETIVOS

El objetivo principal de este estudio es la realización de un análisis bioinformático de los resultados de un array de Illumina en el que se comparó el perfil de expresión génica de muestras de fibroblastos embrionarios transformados WT o KO para Ambra1. Este objetivo general se puede dividir en los siguientes objetivos concretos:

- Desarrollar un código en R que permita analizar los resultados del array.
- Identificar los genes que se encuentran expresados diferencialmente entre las células WT y KO para Ambra1.
- Analizar la expresión de Ambra1 y de genes relacionados con algunos de los procesos previamente identificados como regulados por Ambra1 en el array.
- Analizar los genes expresados diferencialmente entre células WT y KO para Ambra1 utilizando la herramienta Gene Ontology.
- Analizar los genes expresados diferencialmente entre células WT y KO para Ambra1 utilizando la herramienta KEGG.
- Analizar los genes expresados diferencialmente entre células WT y KO para Ambra1 utilizando Reactome.

1- INTRODUCCIÓN

1.1- Autofagia

La autofagia es un proceso de degradación celular que se encuentra altamente conservado en todos los eucariotas. Así, se sabe que la autofagia ejerce un papel clave en el mantenimiento de la homeostasis celular debido a su labor en la degradación de proteínas, orgánulos, macromoléculas y otros componentes celulares.

Existen tres tipos principales de autofagia, macro-autofagia, micro-autofagia, y autofagia mediada por chaperonas (CMA). Aunque existen muchas diferencias entre estos tres tipos de autofagia, todas ellas tienen en común el hecho de que su activación permite la entrega y posterior degradación de diversos componentes celulares “cargos” en el lisosoma ([Parzych KR. And Kilonsky DJ., 2014](#)). En este trabajo se utilizará el término autofagia para referirse a la macro-autofagia.

La autofagia se inicia con la segregación progresiva del material citoplásmico por estructuras de doble membrana conocidas como fagóforos o membranas de aislamiento. Los fagóforos se nuclean del retículo endoplásmico (ER), pero se ha demostrado que otros orgánulos membranosos contribuyen a su alargamiento, incluido el aparato de Golgi, el compartimento intermedio ER-Golgi (ERGIC), la membrana plasmática, las mitocondrias y los endosomas de reciclaje. Los fagóforos completamente sellados, conocidos como autofagosomas, se fusionan con los lisosomas para formar autolisosomas ([Choi, A.M. et al, 2013](#); [Mizushima, N. et al, 2008](#)). Esto promueve la activación de hidrolasas lisosomales y por lo tanto causa la descomposición de la carga autofagosómica. Los productos de estas reacciones catabólicas, aminoácidos, azúcares, ácidos grasos y nucleósidos/nucleótidos; alcanzan el citosol a través de los transportadores de la membrana lisosomal y son reciclados por circuitos anabólicos, ([Fig.1](#)) ([Noboru Mizushima, 2007](#); [Mizushima y Komatsu, 2011](#)).

Es importante señalar que los autofagosomas pueden tomar porciones muy específicas del citoplasma o coger el material intracelular de una manera no-selectiva, dependiendo esencialmente del estímulo inicial ([Stolz et al, 2014](#)).

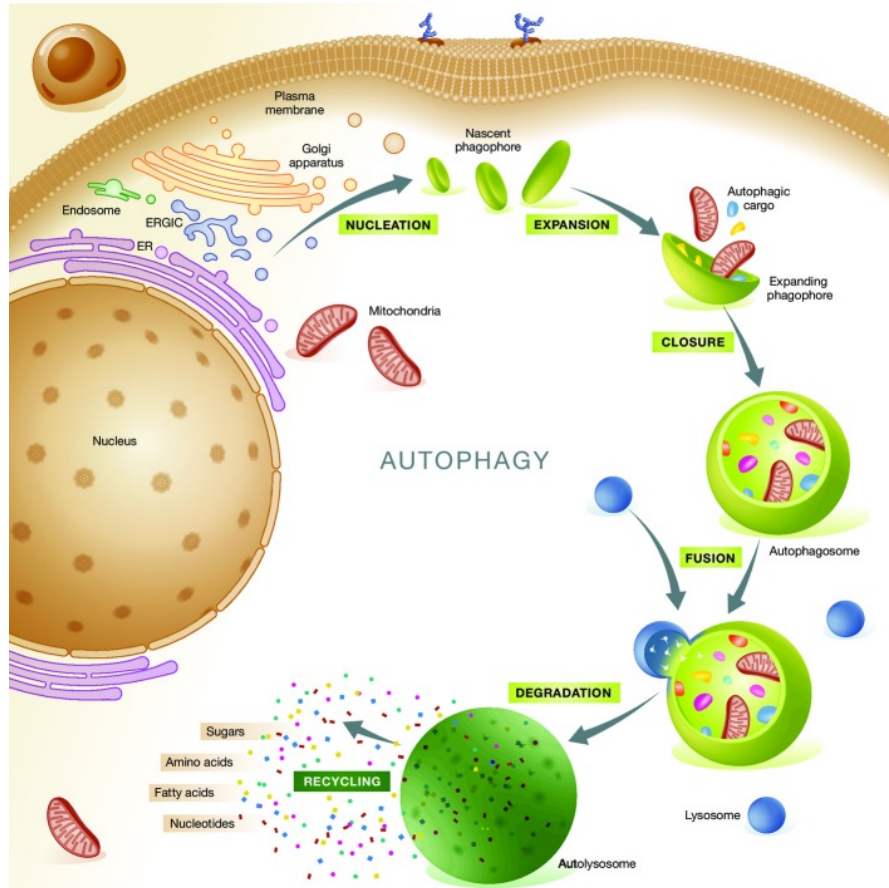


Fig. 1- Esquema del proceso de macro-autofagia ([Galluzzi L., 2015](#))

La autofagia desempeña un papel dual en el control de la supervivencia celular puesto que su activación puede conducir a activar procesos de muerte celular programada o formar parte de un proceso protector ([Choi, A.M. et al, 2013](#); [Mizushima, N. et al, 2008](#)). La autofagia puede tener efecto oncosupresor debido a que está involucrada en el proceso por el cual el sistema inmunológico constantemente elimina células potencialmente tumorogénicas antes de que se ocasionen lesiones malignas ([Ma et al., 2013](#)), o porque constituye una barrera contra procesos degenerativos, incluyendo transformación maligna, que pueden afectar a las células sanas debido a su papel clave en la preservación de la homeostasis intracelular; ([Morselli et al, 2011](#)).

No obstante, la autofagia también mantiene la supervivencia y proliferación de células neoplásicas expuestas a estrés ambiental e intracelular, apoyando por lo tanto, el crecimiento del tumor, la diseminación metastásica y la invasión, en algunos contextos (Kroemer et al, 2010). En cáncer se ha propuesto que la autofagia puede tener un papel importante tanto en el control del crecimiento tumoral como en la respuesta a terapias antitumorales (Verfaillie, T., et al., 2010; White, E., 2012; Galluzzi, L., et al, 2015).

1.2- AMBRA1

AMBRA1 es un factor crucial involucrado en la regulación de la autofagia (Fimia et al., 2007, Fimia y Piacentini, 2010). Su inactivación en vertebrados da como resultado defectos del sistema nervioso en desarrollo y muerte embrionaria (Benato et al., 2013, Cecconi et al., 2008). AMBRA1 se une a BECLIN 1 y regula positivamente la actividad de la lípido quinasa del complejo BECLIN 1 / Vps34 (Di Bartolomeo et al., 2010, Fimia et al., 2007). Se ha demostrado que AMBRA1 regula la actividad y la estabilidad de ULK1, lo que sugiere un papel más amplio de AMBRA1 en la ejecución del programa de autofagia (Nazio et al., 2013).

AMBRA1 tiene la estructura de una proteína intrínsecamente desordenada (IDP), con capacidad de unirse a otros reguladores involucrados en la regulación de varios procesos celulares (Fimia GM, 2013). El desorden intrínseco de Ambra1 favorece la gran plasticidad de esta proteína siendo capaz de coordinar varios procesos intracelulares mediante autofagia (Mei, Y, 2014).

Por ejemplo, la proteína Ambra1 actúa como un regulador positivo en la etapa de iniciación de la autofagia (Fig. 2). Por otra parte, estudios recientes han revelado que Ambra1 también puede coordinar una respuesta de la célula a situaciones de ayuno o de estrés celular (Fimia GM, 2013).

Adicionalmente, Ambra1 es diana de una serie de procesos regulatorios, tales como degradación por ubiquitinación dependiente de culinas (un subtipo de E3 ligasa), rotura por caspasas o fosforilación (Nazio F, 2013). Destaca la importancia de Ambra1 en la regulación de procesos fisiológicos cruciales, tales como metabolismo energético, muerte y división celular. Además, Ambra1 actúa como regulador durante el desarrollo embrionario por lo que su inactivación se ha correlacionado con varias patologías del sistema nervioso así como la carcinogénesis.

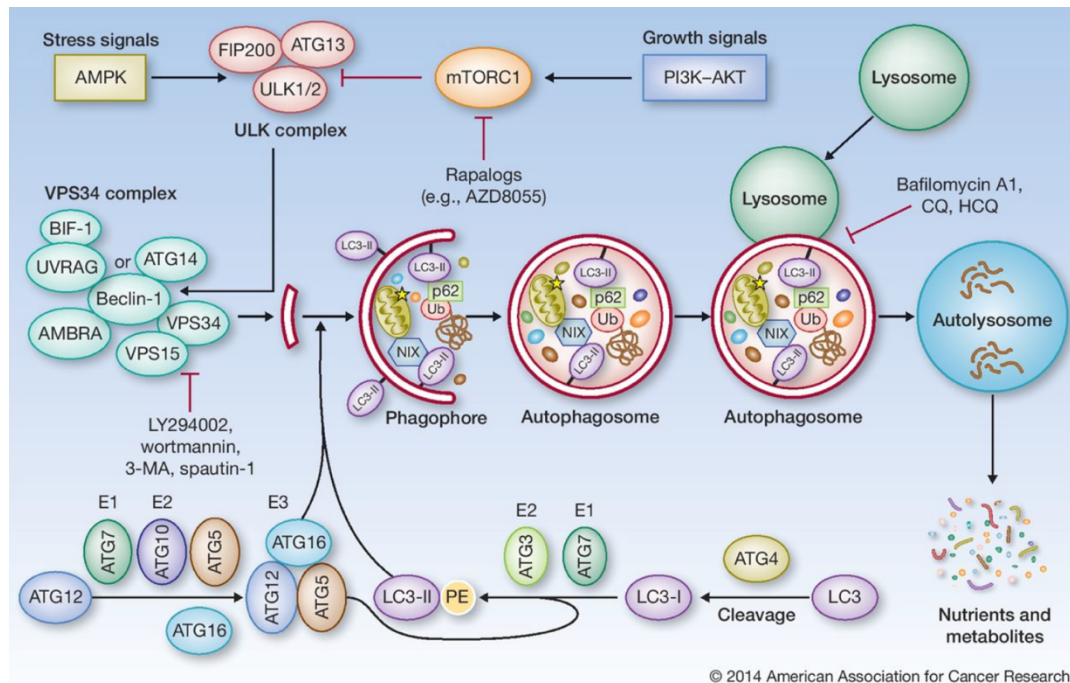


Fig. 2- Proceso de autofagia. Ambra1 favorece la formación del autofagosoma (Cicchini, M, et al., 2014)

La proteína AMBRA1 también regula la proliferación celular, facilitando la desfosforilación y degradación del proto-oncogen C-MYC (Fig. 3). Así, resultados obtenidos por el grupo donde se desarrolló este trabajo en colaboración con el grupo de Francesco Cecconi apoyan la idea de que AMBRA1 actúa como un gen supresor de tumores haplo-insuficiente (Cianfanelli, V. Et al, 2015). Ese papel supresor tumoral de AMBRA1 se debe al menos en parte a su capacidad para regular la interacción entre C-MYC y la fosfatasa PP2A y con ello la fosforilación y estabilidad de C-MYC. Así la inactivación de AMBRA-1 lleva a un aumento de la estabilidad de c-MYC que contribuye a un aumento de la capacidad tumorigénica.

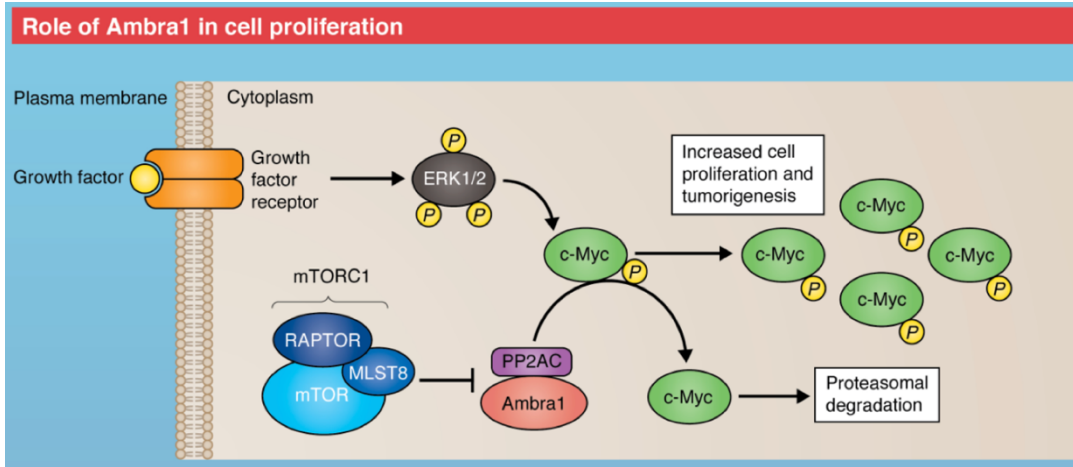


Fig. 3- Ambra1 regula la proliferación celular, (Cianfanelli V., 2015)

1.3- Fibroblastos

Los fibroblastos son células características del tejido, proceden de las células cutáneas del tejido conectivo, en ellos tiene lugar la síntesis de matriz y fibras. Tienen una forma irregular con prolongaciones (Fig.4).

Los fibroblastos activos adquieren tamaño superior a los inactivos, con mayor cantidad de citoplasma y un gran desarrollo de los orgánulos relacionados con la síntesis de los componentes de la matriz extracelular: ribosomas, retículo endoplasmático y complejo de Golgi. Tienen un citoplasma ramificado que rodea un núcleo ovalado y manchado. Los fibroblastos inactivos se llaman fibrocitos, son más pequeños y tienen forma de huso.

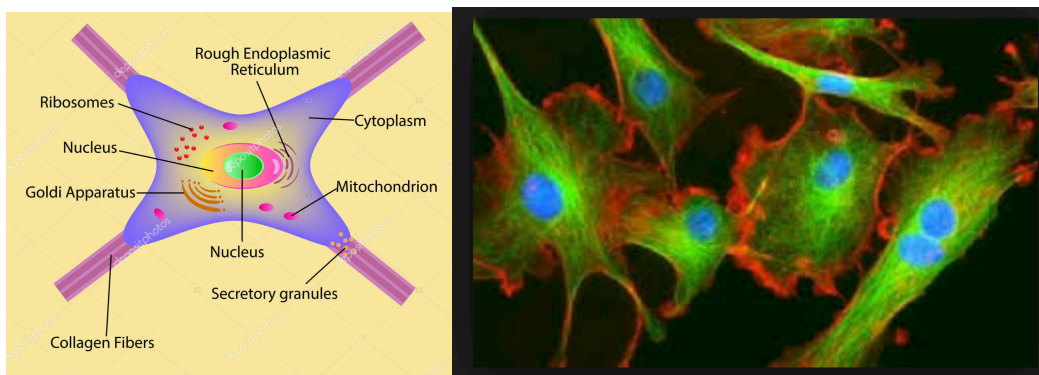


Fig. 4- Esquema e imagen de fibroblasto

Las funciones principales de los fibroblastos son producir la matriz extracelular y el colágeno, éstos conforman el marco estructural de los tejidos en los animales y juegan un papel importante en la reparación de los tejidos y el mantenimiento de la integridad estructural dentro del tejido conectivo. Los fibroblastos secretan los precursores de la matriz extracelular requeridos para la formación de tejido conectivo y diversas fibras.

Los fibroblastos derivan originalmente del mesénquima. Las células epiteliales pueden producir fibroblastos en el proceso denominado transición epitelio-mesénquima (EMT). Por otra parte, los fibroblastos pueden también producir epitelios mediante el proceso conocido como transición mesénquima-epitelio (MET). Este proceso forma parte del desarrollo normal, reparación de tejidos y crecimiento tumoral.

1.4- Transición Epitelio Mesénquima (EMT)

Cuando se desarrolla metástasis, es imprescindible que haya una invasión de las células epiteliales a la lámina basal. Dicha invasión sucede durante el proceso de EMT, produciéndose un conjunto de eventos celulares que permite la transformación de células epiteliales a células migratorias ([Hay ED., 1995](#)).

En el proceso EMT, tiene lugar una ganancia de polaridad, motilidad migratoria y celular de las células epiteliales, así como una pérdida de polaridad apical basal, unión célula-célula y moléculas de adhesión facilitando de esta manera la invasión y metástasis durante la progresión del cáncer ([Lamouille S., et al., 2014](#)), ([Fig.5](#)).

Al iniciarse la EMT se produce el desmontaje del contacto célula-célula epitelial, es decir de las uniones adherentes, uniones estrechas, uniones de tipo desmosoma y gap, lo que se ve acompañado por una pérdida de la polaridad de la célula. El proceso prosigue por la reorganización del citoesqueleto de actina y por la adquisición de motilidad y capacidad invasiva de las células, que forman invadopodios, lamellipodios y filopodios. Asimismo tiene lugar la expresión de metaloproteinasas de matriz (MMPs) que pueden degradar proteínas de la matriz extracelular ([Thiery JP, Sleeman JP., 2006](#); [Kalluri R, Weinberg RA, 2010](#)).

El proceso está regulado por la represión de la expresión de genes que determinan la existencia de un fenotipo epitelial y con la activación de la expresión de genes que regulan la adquisición del fenotipo mesenquimal. Esta compleja transformación involucra por tanto a diversos factores de transcripción y a sus dianas transcripcionales, fundamentalmente E-Cadherina, Catenina, Fibronectina, y Vimentina, que contribuyen a determinar que la célula

adquiera un fenotipo mesenquimal o epitelial. Específicamente, la vía de señalización del Factor de Crecimiento Transformante beta (TGF- β) es crucial para la regulación de la EMT (Thiery JP., 2009).

Las células que han experimentado el proceso EMT pueden revertirlo y volver al estado epitelial a través del proceso de transición mesénquima-epitelio (MET). Tras la difusión, las células cancerosas pueden revertir mediante MET a un estado epitelial, conduciendo a la formación de los carcinomas secundarios con fenotipos similares al tumor primario (Lamouille S., 2014).

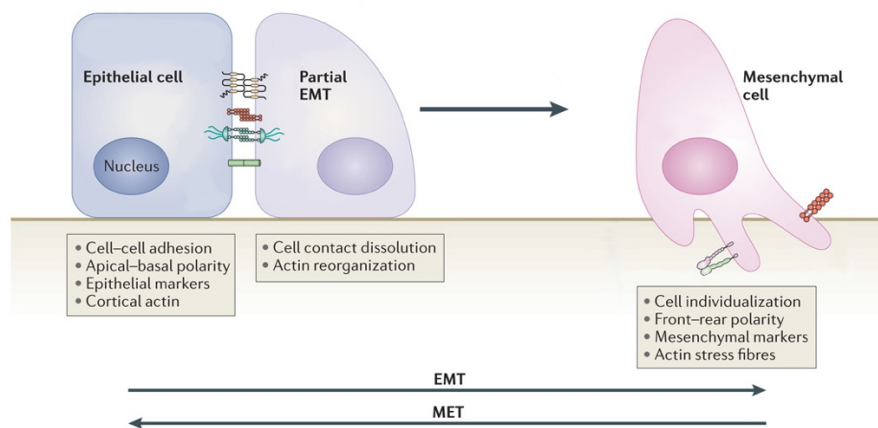


Fig. 5- Primeros pasos de la transición EMT (Lamouille S., 2014)

1.5- GENE ONTOLOGY (GO)

La ontología de genes es un proceso sistematizado que define las funciones de los genes ("términos GO") y cómo estas funciones se relacionan entre sí ("relaciones"), (The Gene Ontology Consortium, 2000). GO se revisa y amplía constantemente a medida que se amplía el conocimiento biológico. Todos los términos de GO tienen un nombre y un identificador único de la forma GO:nnnnnnn, la mayoría con una definición textual, con referencia a la fuente a partir de la cual fue descrito (GO Consortium, 2017). La ontología GO se estructura como un gráfico acíclico dirigido donde cada término tiene relaciones definidas con uno o más términos en el mismo dominio y, a veces, con otros dominios (Fig. 6).

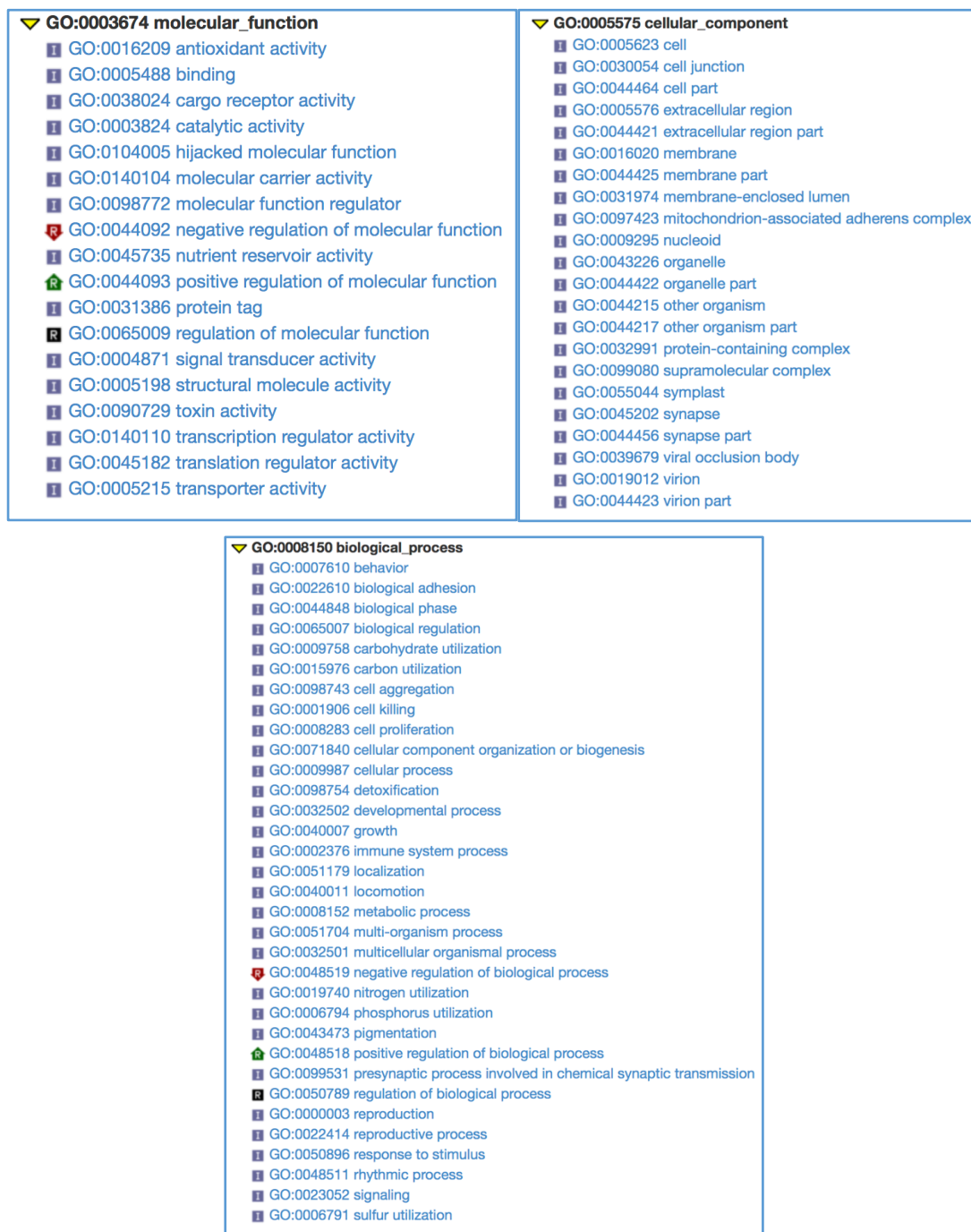


Fig. 6- Esquema de los árboles de GO: MF, CC, BP

GO describe la función con respecto a tres aspectos diferentes; función molecular (MF), componente o localización sub-celular (CC) y proceso biológico (BP). Aunque GO incluye fundamentalmente conceptos que se refieren al nivel sub-celular y celular, abarca también niveles superiores, como los correspondientes a sistemas, órganos y organismo.

La MF describe actividades que ocurren a nivel molecular; sus términos representan a las actividades que llevan a cabo las acciones, sin especificar cuándo, dónde, o en qué contexto ocurren.

El CC se refiere al espacio celular donde se encuentra el producto génico. Un componente celular puede ser una estructura anatómica, como el núcleo celular, el retículo endoplasmático, o una estructura molecular más simple, como un ribosoma o un dímero proteico.

El BP implica generalmente transformaciones químicas o físicas que ocurren por la acción de un conjunto de funciones moleculares organizadas. Los BPs pueden ser de un nivel más elevado o abstracto, como son el “crecimiento celular” o la “transducción de señales”, o de un nivel menor o más específico como son el “metabolismo de pirimidinas” o la “biosíntesis de AMPc” ([Ashburner et al., 2000](#)).

1.6- Kyoto Encyclopedia of Genes and Genomes (KEGG)

KEGG (Enciclopedia de genes y genomas de Kyoto) es una base de datos que integra información funcional genómica, química y sistémica ([Kanehisa, 1996](#)). El objetivo principal del proyecto es asignar significados funcionales a genes y genomas tanto a nivel molecular como superior. Es una representación computarizada del sistema biológico, que consiste en bloques de construcción moleculares de genes y proteínas (información genómica) y sustancias químicas (información química) que se integran con el conocimiento sobre diagramas de cableado molecular de redes de interacción, reacción y relación (información de sistemas). También contiene información sobre enfermedades y medicamentos (información de salud) como perturbaciones al sistema biológico ([Kanehisa, et al, 2017](#))

Actualmente hay cuatro tipos de operaciones de mapeo disponibles en KEGG; las funciones de nivel molecular se almacenan en la base de datos de KO (KEGG Orthology), donde cada KO se define como un ortólogo funcional de genes y proteínas. Las funciones de nivel superior están representadas por redes de interacciones moleculares, reacciones

y relaciones en forma de mapas de rutas KEGG, jerarquías BRITE y módulos KEGG (Fig. 7).

Como resultado, KEGG se usa ampliamente como una base de conocimiento de referencia para la integración e interpretación de conjuntos de datos a gran escala generados por la secuenciación del genoma y otras tecnologías experimentales de alto rendimiento.

Category	Database name	Content	Release
Systems Information	KEGG PATHWAY	KEGG pathway maps	1995
	KEGG BRITE	BRITE functional hierarchies and tables	2005
	KEGG MODULE	KEGG modules	2006
Genomic Information	KEGG ORTHOLOGY (KO)	KO groups for functional orthologs	2002
	KEGG GENOME	KEGG organisms (complete genomes)	2000
	KEGG GENES	Genes and proteins	1995
	KEGG SSDB	Sequence similarity among GENES entries	2001
Chemical Information	KEGG COMPOUND	Metabolites and other small molecules	1995
	KEGG GLYCAN	Glycans	2003
	KEGG REACTION / RCLASS	Biochemical reactions and reaction class	1998/2010
	KEGG ENZYME	Enzyme nomenclature	1995
Health Information	KEGG NETWORK / VARIANT	Disease-related network and gene variants	2017
	KEGG DISEASE	Human diseases	2008
	KEGG DRUG / DGROUP	Drugs and drug groups	2005/2014
	KEGG ENVIRON	Crude drugs and health-related substances	2010
	JAPIC	Japanese drug labels	2007
	DailyMed	FDA drug labels (links only)	2012

Fig. 7- Esquema de información en KEGG (Tanabe M., et al, 2018)

1.7- REACTOME

REACTOME es una base de datos de código abierto, acceso abierto, curada manualmente y revisada por pares. La base de datos proporciona herramientas de bioinformática para la visualización, interpretación y análisis del conocimiento de las vías para apoyar la investigación básica y clínica, biología de sistemas, análisis del genoma entre otras (Fabregat A., et al, 2018).

Contiene un conjunto de recursos en línea que utilizan las características de los medios electrónicos para organizar la información de la ruta biológica de manera que proporciona un acceso más eficiente y permite nuevas formas de análisis (Fig. 8).

Es una base de datos relacional de moléculas metabólicas y de señalización

organizadas en vías y procesos biológicos. La unidad central del modelo de datos es la reacción. Las entidades (ácidos nucleicos, proteínas, complejos, vacunas, terapias anticancerosas y moléculas pequeñas) que participan en las reacciones forman una red de interacciones biológicas y se agrupan en vías. Los ejemplos de vías biológicas en Reactome incluyen el metabolismo intermediario clásico, la señalización, la regulación transcripcional, la apoptosis y la enfermedad.

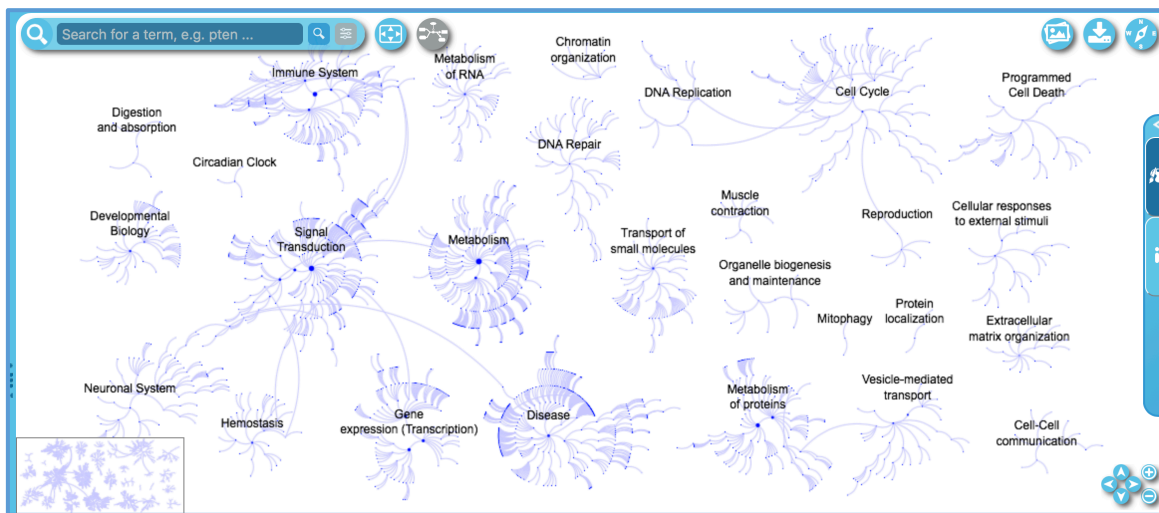


Fig. 8- Herramienta de visualización y análisis de datos REACTOME

1.8- Gene Set Enrichment Analysis (GSEA)

El análisis de enriquecimiento de conjuntos de genes (GSEA) es un método computacional que determina si un conjunto de genes definidos a priori muestra diferencias estadísticamente significativas y concordantes entre dos estados biológicos ([Subramanian, Tamayo, et al., 2005](#); [Mootha, Lindgren, et al et al, 2003](#)). El método se basa en concentrar el análisis en grupos de genes que comparten una función biológica, una ubicación cromosómica o una regulación.

El método GSEA está incorporado en un paquete de software abierto, con una base de datos inicial de 1,325 conjuntos de genes definidos biológicamente.

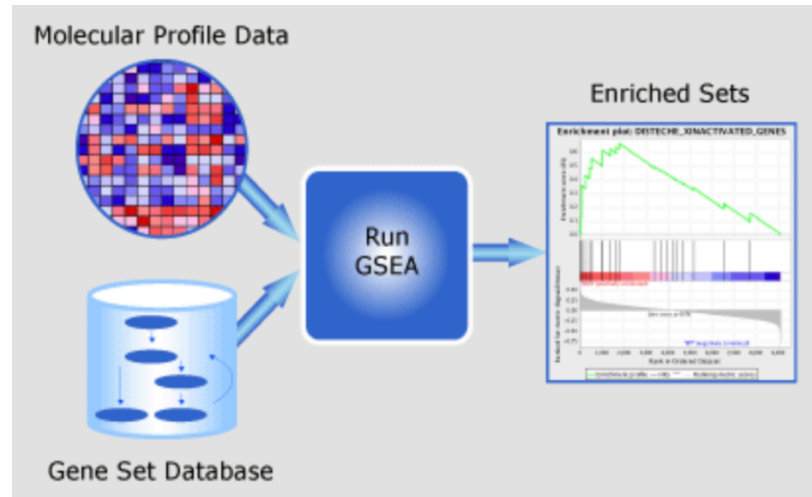


Fig. 9- Esquema funcionamiento GSEA

2- MATERIALES Y MÉTODOS

2.1- R

R es un lenguaje y entorno de programación para la manipulación y el análisis de datos estadísticos y gráficos. Se trata de una implementación de dominio público del lenguaje estadístico S, base del software de código abierto S-PLUS. R es un lenguaje de programación de propósito general orientado a objetos, por lo que se puede utilizar para automatizar los análisis y crear nuevas funciones que amplíen las características del lenguaje existente (Matloff N., 2011). El lenguaje de programación R se distribuye bajo Windows, MacOS y Linux e incluye consola, editor que soporta la ejecución de código directo, así como herramientas para la gestión de gráficos, (<https://www.rstudio.com/products/RStudio/>). Además de proporcionar operaciones estadísticas R cuenta con un conjunto amplio de funciones de análisis de datos, con funciones muy flexibles para realizar gráficos.

Para el desarrollo del presente proyecto desarrollamos un programa en R para realizar el análisis de un array de Illumina del cual obtuvimos un listado de genes expresados diferencialmente (Fig. 10).

Para ello utilizamos BioConductor y varias de sus librerías.

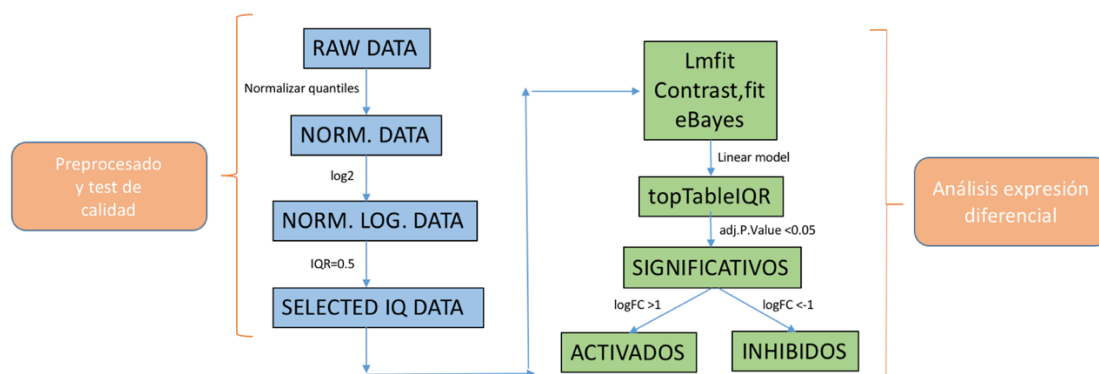


Fig. 10- Flujo de trabajo realizado con los datos del array de Illumina para obtener listado de genes expresados diferencialmente DEGs

Una vez obtenidos los genes expresados diferencialmente realizamos un script para analizar la ontología de genes, asignar significados funcionales a genes, relacionar moléculas metabólicas y de señalización organizadas en vías y procesos biológicos y determinar si un conjunto de genes definidos a priori muestra diferencias

estadísticamente significativas y concordantes entre dos estados biológicos. Para ello realizamos el flujo de trabajo que se observa en la (Fig. 11).

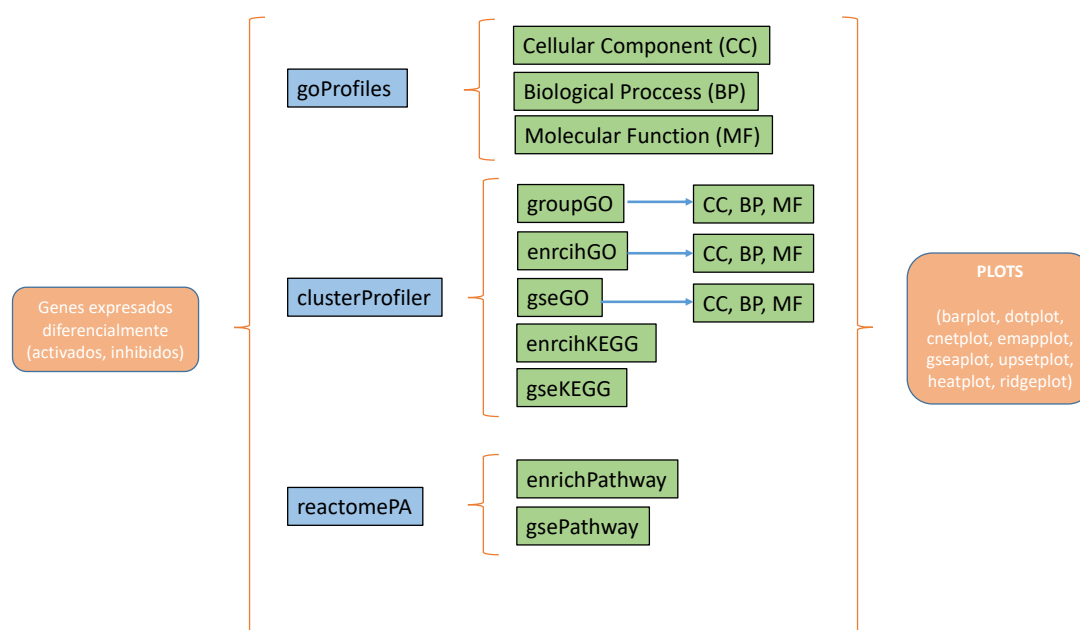


Fig. 11- Flujo de trabajo realizado a partir del listado de genes expresados diferencialmente para análisis GO, KEGG, REACTOME, GSEA.

2.2- Bioconductor

Bioconductor es un proyecto de software de código abierto y desarrollo abierto para el análisis y la comprensión de datos de alto rendimiento en genómica y biología molecular (<https://www.bioconductor.org>). Utiliza el lenguaje de programación estadística R, es de código abierto y desarrollo abierto. Tiene dos lanzamientos cada año y 1473 paquetes de software. Los paquetes cubren una gama de aplicaciones bioinformáticas y estadísticas y se someten a una revisión inicial formal y pruebas automatizadas continuas. (Huber et al., 2015; Gentleman et al., 2004).

Bioconductor proporciona acceso generalizado a una amplia gama de potentes métodos estadísticos y gráficos para el análisis de datos genómicos, facilita la inclusión de metadatos biológicos en el análisis de datos genómicos y aporta una plataforma de software común que permite el rápido desarrollo e implementación de software extensible, escalable e interoperable. Los paquetes de análisis abordan los flujos de trabajo para el análisis de matrices de oligonucleótidos, análisis de secuencia, citometría de flujo y otros datos genómicos de alto rendimiento.

La mayoría de los componentes de Bioconductor se distribuyen como paquetes de R. El alcance funcional de los paquetes de Bioconductor incluye el análisis de microarrays de ADN, secuencia, flujo, SNP y otros datos.

2.2.1- Limma

Linear Models for Microarray Data (LIMMA), es una biblioteca para el análisis de datos de microarrays de expresión génica; se utilizan especialmente modelos lineales para analizar experimentos diseñados y se evalúa la expresión diferencial. (<http://bioconductor.org/packages/release/bioc/html/limma.html>)

LIMMA proporciona la capacidad de analizar comparaciones entre muchos ARN simultáneamente en experimentos diseñados arbitrariamente. Los métodos bayesianos empíricos se utilizan para proporcionar resultados estables incluso cuando el número de matrices es pequeño. El modelo lineal y las funciones de expresión diferencial se aplican a todas las tecnologías de expresión génica, incluidas microarrays, RNA-seq y PCR cuantitativa ([Ritchie ME, et al., 2015](#)).

En el presente estudio se utilizó Limma para analizar los datos crudos provenientes del array de DNA realizado a partir de fibroblastos embrionarios derivados de ratones WT o KO para Ambra1.

2.2.2- gplots

La biblioteca gplots contiene varias herramientas de programación R para hacer gráficos, entre las que caben destacar; para gráficos estándar 'barplot2', 'boxplot2', 'heatmap.2', 'smartlegend'; para cálculo y trazado de resúmenes de datos bidimensionales 'ci2d', 'hist2d'; el interfaz para fórmulas de la función 'stats :: lowess', 'lowess', ([Warnes GR., et al., 2016](#)), (<https://CRAN.R-project.org/package=gplots>).

2.2.3- ggplot2

ggplot2 es un paquete de R para realizar gráficos estadísticos o de datos. Tiene una gramática, basada en la Gramática de los gráficos ([Wilkinson, 2005](#)), está compuesta por un conjunto de elementos independientes que se pueden organizar de muchas maneras diferentes. Esto hace que ggplot2 sea muy potente, ya que no está limitado a un conjunto de gráficos pre-especificados, sino que puede crear nuevos

gráficos que se adaptan con precisión a cada problema (Wickham H., and Chang W., 2016), (<https://CRAN.R-project.org/package=ggplot2>).

Está diseñado para trabajar en capas, comenzando con una capa que muestra los datos brutos y luego agrega capas de anotaciones y resúmenes estadísticos.

2.2.4- illuminaMousev2.db

Base de datos de Illumina para realizar anotación de los genes a partir de los identificadores del fabricante. Los datos de anotación de Illumina Mousev2 (chip illuminaMousev2) se ensamblaron utilizando datos de repositorios públicos. Las sondas Illumina con un puntaje de “Sin coincidencia” o “Mala calidad” se eliminaron antes de la anotación, (Dunning M, Lynch A and Eldridge M., 2015). El nombre representa el símbolo del gen y el vector contiene todos los identificadores del fabricante que se encuentran para ese símbolo. (<http://bioconductor.org/packages/illuminaMousev2.db/>)

illuminaMousev2SYMBOL, cada identificador del fabricante se asigna a una abreviatura para el gen correspondiente. Se informa una NA si no se conoce una abreviatura para un gen determinado. Los símbolos suelen constar de 3 letras que definen un solo gen (ABC) o genes múltiples (ABC1, ABC2, ABC3). Los símbolos genéticos se pueden usar como palabras clave para consultar bases de datos públicas como Entrez Gene.

2.2.5- goProfiles

Realiza análisis basados en ontología de genes para conjuntos de genes u otro tipo de identificadores biológicos que pueden anotarse en la ontología de genes. La creciente disponibilidad de datos de alto rendimiento, cantidades masivas de datos de biología molecular que surgen de diferentes análisis, como la expresión de genes o microarrays, conlleva la necesidad de métodos para resumir la información disponible.

goProfiles utiliza GeneOntology (GO), para construir perfiles funcionales que caracterizan un conjunto de genes o proteínas utilizando la distribución de sus anotaciones en la base de datos; proporciona métodos para comparar los perfiles y desarrolla procedimientos independientes para evaluar esta comparación (Sanchez A, Ocana J, Salicru M, 2018), (<http://bioconductor.org/packages/goProfiles/>). goProfiles fue utilizado para la construcción de los perfiles funcionales de los datos de microarrays incluidos en el presente TFM.

2.2.6- clusterProfiler

El paquete clusterProfiler implementa métodos para analizar y visualizar perfiles funcionales de coordenadas genómicas, grupos de genes y genes individuales. Es un paquete de R / Bioconductor que automatiza el proceso de clasificación de términos biológicos y el análisis de enriquecimiento de grupos de genes (Yu G., et al, 2012) , (<http://bioconductor.org/packages/clusterProfiler/>). El módulo de análisis y el módulo de visualización se combinan en un flujo de trabajo reutilizable. Admite tres especies, incluidos humanos, ratones y levadura aunque los métodos proporcionados pueden extenderse fácilmente a otras especies y ontologías. Incorpora el conocimiento biológico (GO) y la Enciclopedia de Genes y Genomas (KEGG), para identificar los temas biológicos predominantes de una colección de genes. Las ontologías y vías que permite manejar el paquete son: ontología de la enfermedad (Disease Ontology, DOSE), Red de genes de cáncer (DOSE), DisGeNET (DOSE), soporta muchas especies con la anotación GO (AnnotationHub), KEGG Pathway and Module con los últimos datos en línea (http://www.genome.jp/kegg/catalog/org_list.html), Reactome Pathway (ReactomePA), DAVID (RDAVIDWebService), entre otras.

2.2.7- ReactomePA

El paquete Reactome es una base de datos de anotación de ruta curada manualmente para obtener rutas biológicas de alto orden a partir de datos de alto rendimiento. Es un paquete de R / Bioconductor que proporciona análisis de enriquecimiento, que incluye pruebas hipergeométricas y análisis de enriquecimiento de conjuntos de genes. Se puede aplicar un análisis funcional a la coordinación genómica obtenida de un experimento de secuenciación para analizar la importancia funcional de los loci genómicos, incluidos los elementos reguladores en cis y las regiones no codificantes. Permite realizar comparaciones entre diferentes experimentos, (Yu G, He Q., 2016) , (<http://bioconductor.org/packages/ReactomePA/>).

2.2.8- DOSE

La ontología de la enfermedad (Disease Ontology, DO) anota los genes humanos en el contexto de la enfermedad. DOSE es un paquete de R que proporciona cálculos de similitud semántica entre los términos de DO y los genes. El paquete

implementa los análisis de enriquecimiento que incluyen el modelo hipergeométrico y el análisis de enriquecimiento de conjuntos de genes para ayudar a encontrar asociaciones de enfermedades de datos biológicos de alto rendimiento. Permite realizar comparaciones entre grupos de genes, (Yu G, He Q., 2016) , (<http://bioconductor.org/packages/ReactomePA/>).

DOSE proporciona cinco métodos para medir la similitud semántica entre los términos y genes de DO. Los análisis de enriquecimiento se basan en la ontología de la enfermedad (DO), Network of Cancer Gene (NCG) y DisGeNET. El paquete cuenta con varios métodos de visualización para ayudar a interpretar los resultados semánticos y de enriquecimiento.

2.2.9- topGO

El paquete topGO está diseñado para facilitar el análisis de enriquecimiento semiautomático para los términos de Ontología Genética (GO). Proporciona la interpretación y visualización del análisis de enriquecimiento de los términos GO procedentes de resultados de mediciones de expresión génica, correlación de genes o análisis de expresión diferencial, (Alexa A, Rahnenfuhrer J, 2018), (<http://bioconductor.org/packages/topGO/>)

2.2.10- enrichplot

El paquete enrichplot implementa varios métodos de visualización para ayudar a interpretar los resultados obtenidos de enriquecimiento. Permite la visualización de resultados de enriquecimiento obtenidos de DOSE (Yu et al. 2015), clusterProfiler (Yu et al. 2012) y ReactomePA (Yu and He 2016). Además es compatible con los análisis de representación (Over Representation Analysis, ORA) y el análisis de enriquecimiento de conjuntos de genes (GSEA), (Yu G., 2018), (<http://bioconductor.org/packages/enrichplot/>).

2.2.11- GO.db

Conjunto de mapas de anotaciones que describen la ontología génica completa ensamblada utilizando datos de GO. Proporciona información detallada sobre la última versión de las ontologías genéticas. El paquete se actualiza bianualmente, (Carlson M., 2018), (<http://bioconductor.org/packages/GO.db/>).

2.2.12- KEGG.db

Conjunto de mapas de anotación para KEGG ensamblados utilizando datos de KEGG. Proporciona información detallada de las bases de datos de la ruta KEGG, (Carlson M., 2016), (<http://bioconductor.org/packages/KEGG.db/>).

Base de datos para comprender funciones y utilidades de alto nivel del sistema biológico a partir de información de nivel molecular, especialmente conjuntos de datos moleculares a gran escala generados por la secuenciación del genoma y otros procesos de alto rendimiento.

2.2.13- KEGGREST

El paquete KEGGREST proporciona una interfaz de cliente al servidor REST de KEGG. Basado en KEGGSOAP pero más simple y potente. KEGGREST permite el acceso a la API REST de KEGG. La API de REST de KEGG se basa en operaciones simples: keggInfo (), keggList (), keggFind (), keggGet (), keggConv y keggLink (), (Tenenbaum D.,2018), (<http://bioconductor.org/packages/KEGGREST/>).

2.2.14- reactome.db

Conjunto de mapas de anotación para reactome ensamblados utilizando datos de reactome. El paquete proporciona información detallada sobre la última versión de la base de datos Reactome. Se actualiza bianualmente. (Ligtenberg W., 2018), (<http://bioconductor.org/packages/reactome.db/>)

3- RESULTADOS

3.1- Resultados experimentales previos

Diversos estudios desarrollados por el grupo donde se ha llevado a cabo este TFM tanto in vitro como en modelos animales apoyan la idea de que una pérdida o disminución en la expresión de la proteína Ambra1 favorece el desarrollo de carcinoma escamoso de piel (SCC). Además, algunas observaciones obtenidas a partir de muestras de pacientes sugieren que una menor expresión de Ambra1 se correlaciona con un fenotipo más agresivo de las lesiones y una progresión acelerada de la enfermedad. Estas observaciones apoyan la idea de que Ambra1 podría ser un buen biomarcador cuyo nivel de expresión sería capaz de contribuir a determinar una mayor propensión a desarrollar metástasis. Teniendo en cuenta que histopatológicamente las lesiones malignas y benignas son casi idénticas, de confirmarse estas observaciones resultarían de gran utilidad para mejorar el diagnóstico y tratamiento del SCC.

Teniendo en cuenta todo lo anterior, en el presente trabajo analizamos un array de Illumina cuyas muestras se obtuvieron a partir de fibroblastos embrionarios (día E13.5) derivados de ratones WT o KO para Ambra1 y posteriormente transformados con la forma constitutivamente activa de Ras (V12) y la proteína viral E1A. Una vez seleccionadas las células KO fueron transfectadas con un vector que expresaba la forma WT de Ambra1 o con un plásmido que codificaba para una forma mutante de la proteína (PxP) en la que se había eliminado el dominio implicado en la interacción con PP2A. El RNA se obtuvo utilizando un kit de Qiagen de acuerdo a las instrucciones del fabricante.

En un trabajo previo ([Mendiburu-Eliçabe M., 2018](#)), realizamos la detección de genes expresados diferencialmente (DEGs), que cambian su expresión en función de variables de interés, realizamos el ajuste al modelo lineal de los datos. Calculamos las estadísticas t moderadas de Bayes empíricas utilizando el paquete limma de Bioconductor. El ajuste de los valores de p se realizó mediante la determinación de las tasas de falso descubrimiento (FDR) mediante el procedimiento de Benjamini- Hochberg. Efectuamos test de calidad de los datos, boxplot, clustering de las muestras.

Cuando realizamos la comparación entre los grupos wt y KO encontramos que existen 4574 genes que se expresan de forma diferencial en los grupos, que cumplen $\text{adj.P.val} < 0.05$. De éstos, 375 están activados ($\text{FDR} > 1$) y 504 están inhibidos ($\text{FDR} < -1$) ([Figs. 12, 13, 14](#)).

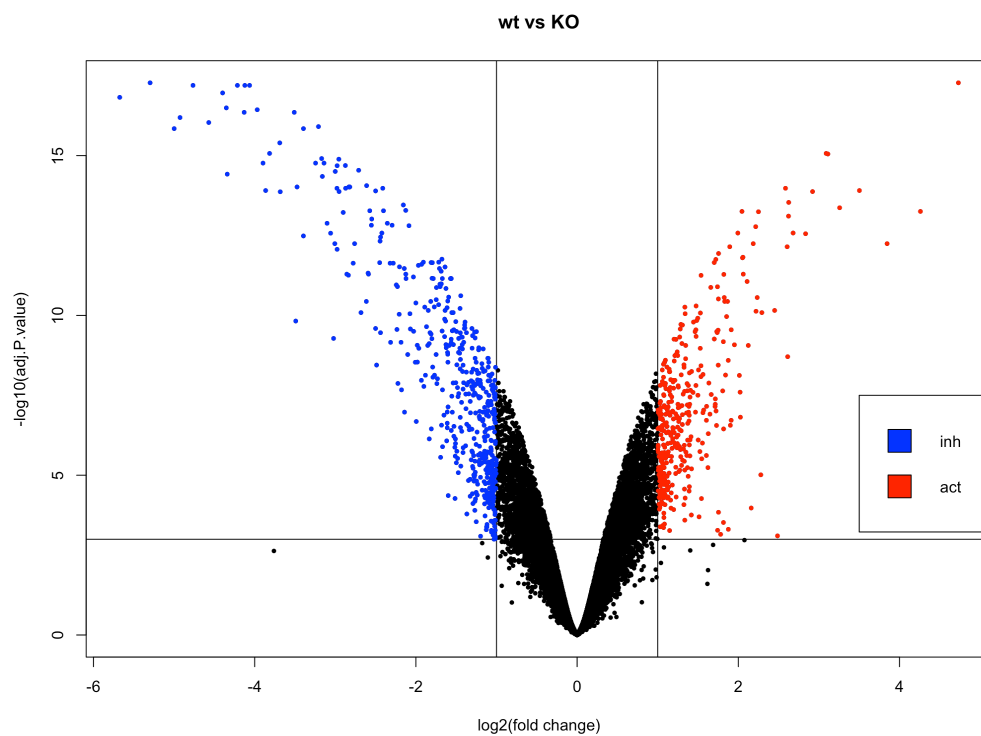


Fig. 12- Volcanoplot wt-KO, genes activados (rojo), genes inhibidos (azul)

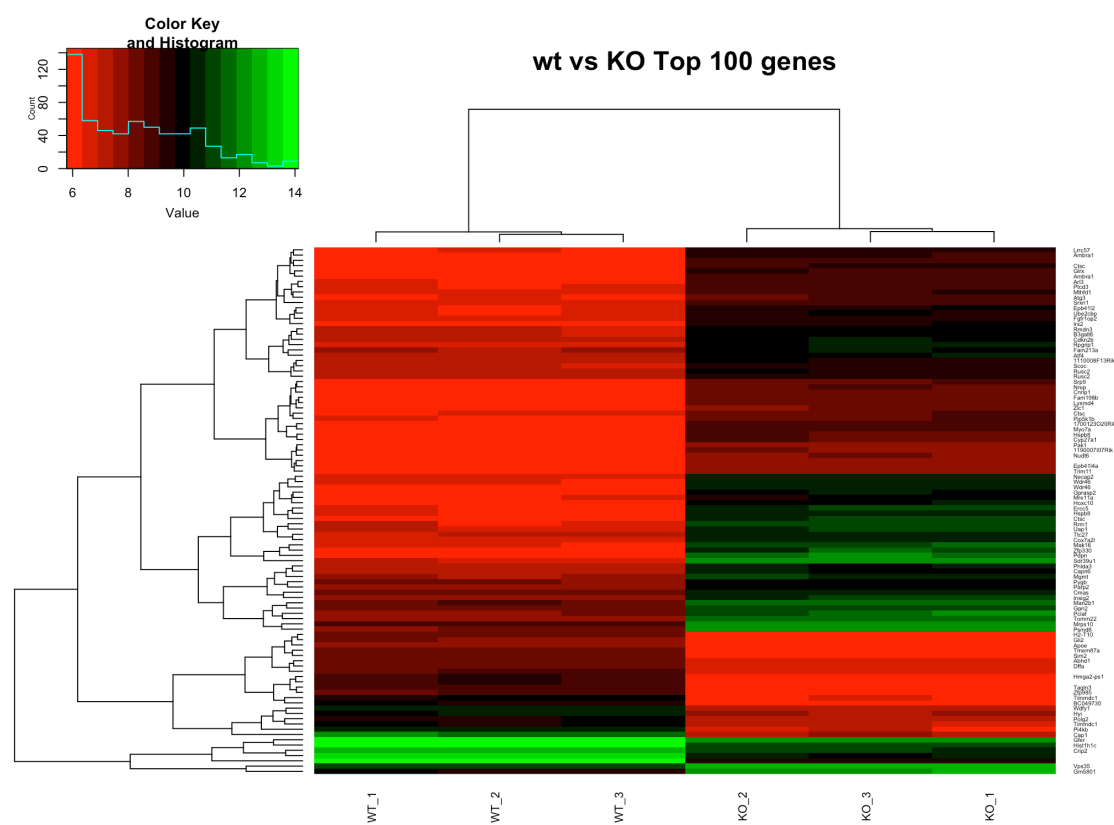


Fig. 13- Heatmap wt-KO, top 100 genes

Genes activados:

NA	Gfer	Sirpa	Car12	Zmat5	Dhh	Hist2h2aa1	Srprb
NA	Crip2	Nicn1	Timp1	Ggnbp1	Ercc8	Nubp2	Btbd6
Wdfy1	NA	Ttl	Zfp862-ps	Fahd2a	Golt1b	Tmem87a	Smagp
Tagln3	Dffa	Galnt11	Il11	Elf5a	Hs3st3b1	Nle1	Aebp1
BC049730	Apoe	NA	Ptpn6	Hs3st3a1	Il17rc	Sptbn2	Gprc5c
Hmga2-ps1	Cdh15	NA	Nubp2	Nek3	Elf5a	Zfp691	Trpv2
Hist1h1c	Notch3	Lpar2	Mfap2	NA	Bhlhe40	2410006H16Rik	Fahd2a
Timmdc1	Gm8909	Ino80c	Prickle3	NA	Btbd6	Bckdhb	Hist1h2bf
Cap1	Ndufb10	Kyat3	BC028528	Vrk2	Chchd10	Patz1	Abhd1
H2-T10	NA	Adat2	NA	Stc2	Gfpt2	Adora2b	NA
Sim2	A930005H10Rik	Slco2a1	Abcc10	Usf1	Etv2	Hdh3	Ubox5
NA	Sct	NA	Emp3	Mindy1	Slco2a1	Il17rc	Arhgdib
Polg2	Ppfbp1	Timp1	Clec16a	Hist2h2aa1	Lypd3	Fam20b	Alad
Zfp985	Rps6kl1	NA	Ino80c	Hist1h3a	4930555G01Rik	Pimreg	Sys1
Abhd1	Erdr1	Hist1h1c	Prodh	Zfp984	NA	Clns1a	Mrps27
Timmdc1	Hif1an	Atp10d	Mindy1	Pkd2l1	Atox1	Prodh	Espn
Pi4kb	Hsp90ab1	Nek3	Cryz	Nipal3	NA	Pemt	Ciapi1
Tmem87a	Atf3	Pi4kb	NA	Sirt5	Ubac2	Itgb3bp	Dusp2
Gli2	Emp3	Hebp1	Plxna2	Mnd1	Eno1b	Upp1	Zkscan17
Hyl	Pdlim4	Nacc1	Nek3	Glo1	Sigirr	NA	Acbd4

Genes inhibidos:

Sdr39u1	Lrrc57	Pclaf	Vps35	Cwf19l1	2310039H08Rik	St6galnac2	Apex1
Tomm22	Ambra1	Fam198b	NA	Dlgap5	Picalm	NA	Nat8f4
Wdr46	Uap1	Parp2	Epb4114a	Fut10	Slc35g2	Gas6	Parp8
Psmid8	Capn6	Scoc	Trim11	D830030K20Rik	Rhoj	Aldh18a1	A230050P20Rik
Mak16	Ambra1	Hspb8	Atg3	Fggy	Srpr	Acsf6	Plpbb
Wdr46	1700123O20Rik	Cdkn2b	Nudt6	Slc35d2	Cd3eap	Ociad1	BC034090
Pdpn	B3gal6	Rusc2	Cnrip1	Psmid12	Cyp27a1	Cavin1	Gcnt1
Hspb8	Rrm1	Cmas	Atf4	Mapk13	Rpp14	Ctgef	Nrros
Necap2	Man2b1	Ube2cbp	Rusc2	4933427D14Rik	B3gnt3	Rlok3	NA
Gprasp2	Arl3	Srp9	Pygb	Hif1an	Ccndbp1	Ehd4	Bbs7
Ttc27	NA	1110008F13Rik	Zic1	Apbb1ip	Amotl1	Map9	Leo1
Ctsc	Gpn2	Lysmd4	Ctsc	Zfp869	Stat4	Pip4k2a	Npc1
Ercc5	Gm5801	Nrep	Pak1	Cbwd1	Fam213a	Apoa2	Zbtb40
Rmdn3	Rpgrip1	Epb4112	Pip5k1b	Zfp35	Tceal6	Cdc37l1	Pla2g12a
Mrps10	Insig2	Irx2	1190007I07Rik	Rp9	Med23	Slc16a13	Tubgcp4
Zfp330	Glrx	Cyp27a1	Tmem82	Acer2	Pop4	Pi4k2b	Grb10
Cox7a2l	Srxn1	Mthfd1	Gnaq	Emp2	Mdk	Il6st	Rpl7l1
Hoxc10	Mre11a	Mgmt	Dusp16	Bag2	Cyp27a1	Ndufb10	Mid2
Ctsc	Ptcd3	Fam213a	Fgfbp3	Mttr2	Mpg	Apbb1ip	Nxn
Myo7a	Fgfr1op2	Phlda3	Adgra3	Mapk13	Fastkd3	Srp9	Ssh3

Fig. 14- Listado de genes activados (panel superior), listado de genes inhibidos (panel inferior)

3.2- Análisis del efecto de la inhibición genética de Ambra1

Llevamos a cabo el análisis bioinformático del efecto de la inhibición genética de Ambra1 en fibroblastos embrionarios transformados. Para ello realizamos diferentes análisis de nuestros DEGs con R utilizando distintos paquetes especializados en proceso de clasificación e interpretación de términos biológicos y el análisis de enriquecimiento de grupos de genes ([Sanchez A., 2018](#); [Yu G, 2012, 2016](#)). Analizamos tanto Gene Ontology (GO), para desarrollar una representación computacional de cómo los genes codifican funciones biológicas a nivel molecular, celular y del sistema tisular; como KEGG (Kyoto Encyclopedia of Genes and Genomes), para comprender las funciones de alto nivel y las utilidades del sistema biológico, como la célula, el organismo y el ecosistema, a partir de información genómica y de nivel molecular. También contiene información sobre enfermedades y medicamentos (información sobre la salud) como perturbaciones para el sistema biológico. Analizamos también el Gene Set Enrichment Analysis (GSEA), para realizar un análisis de enriquecimiento de conjuntos de genes (GSEA) (también análisis de enriquecimiento funcional) para identificar clases de genes o proteínas que están representados en exceso en un gran conjunto de genes o proteínas, y puede tener una asociación con fenotipos de enfermedades.

3.2.1- GO análisis de los genes expresados diferencialmente

Utilizamos los paquetes de bioconductor, goProfiles y clusterProfiler para identificar las categorías GO representadas. Realizamos tanto el análisis de componente celular (CC), como el proceso biológico (BP) y la función molecular (MF) de los DEGs, activados e inhibidos de forma independiente.

3.2.1.1- goProfiles

Realizamos el análisis basado en ontología de genes (GO) para el conjunto de genes expresados diferencialmente. Empleamos la función **basicProfile** para calcular el perfil funcional básico de la lista dada de genes, para un nivel determinado del GO, en nuestro caso utilizamos level = 2. Y para una ontología determinada, CC, BP ó MF:

```
bP_act_CC <- basicProfile(geneID_act_entrez[,2], idType = "Entrez", onto = "CC", level = 2, orgPackage="org.Mm.eg.db",  
ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
```

El análisis de GO señala que los DEGs activados están enriquecidos en los componentes celulares (CC) siguientes: orgánulos, membrana, complejo macromolecular y región extracelular ([Fig. 15](#)). Los procesos biológicos (BP) más abundantes para los DEGs activados son: regulación de proceso biológico (negativo y

positivo), organización componente celular o biogénesis, proceso de desarrollo, respuesta a estímulo entre otros (Fig. 16). En cuanto a la función molecular (MF), los DEGs activados están enriquecidos principalmente en binding y actividad catalítica además de actividad reguladora de transcripción y regulación de la función molecular (Fig. 17).

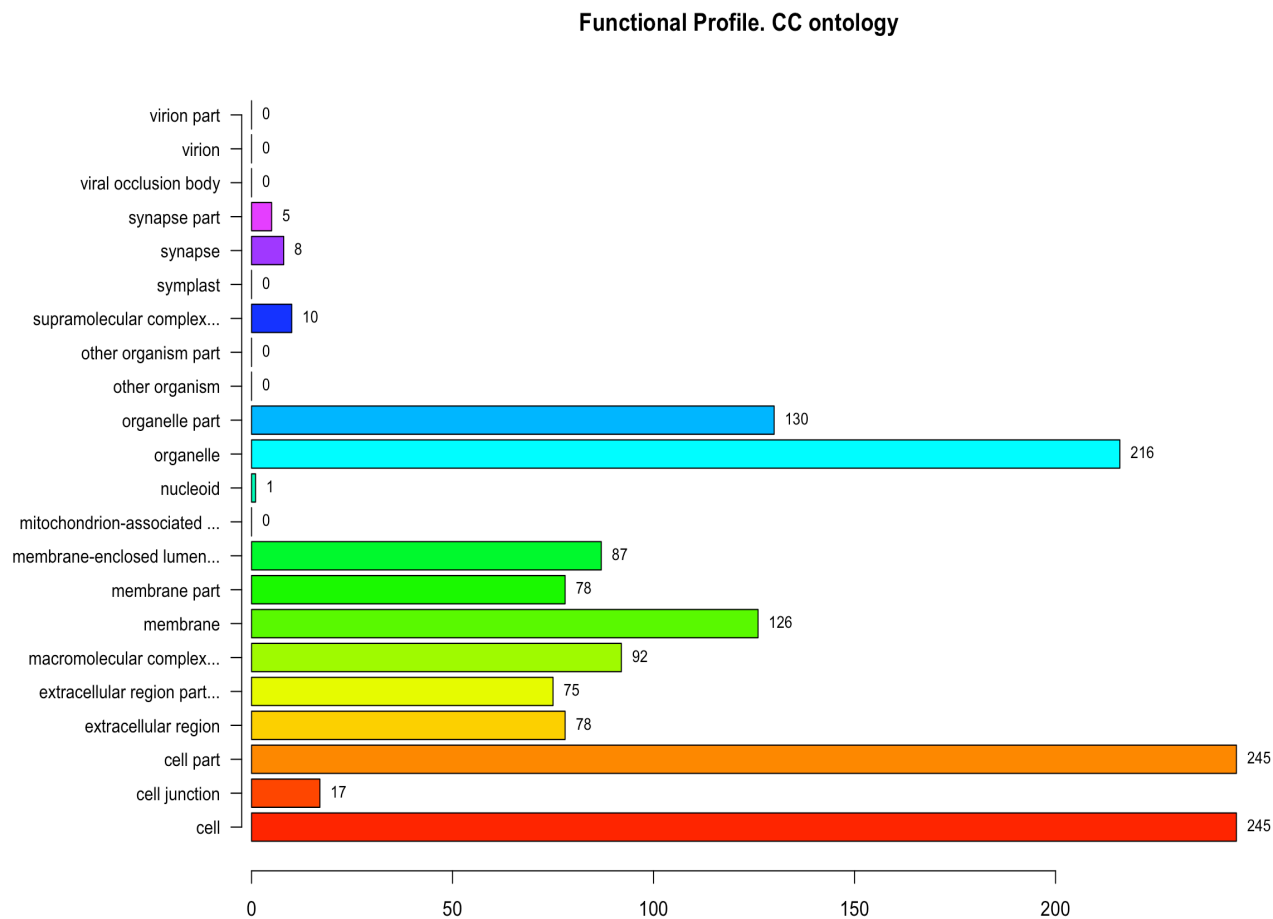


Fig. 15- Barplot de DEGs activados, GO componente celular (CC).

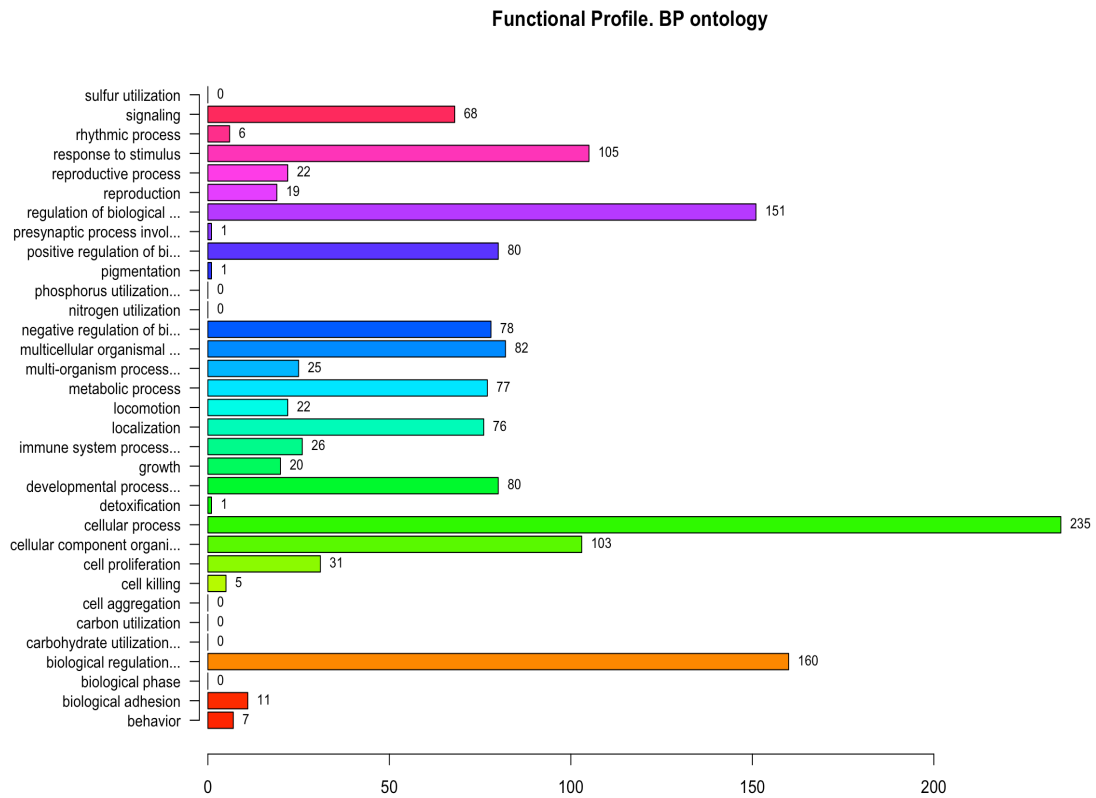


Fig. 16- Barplot de DEGs activados, GO proceso biológico (BP).

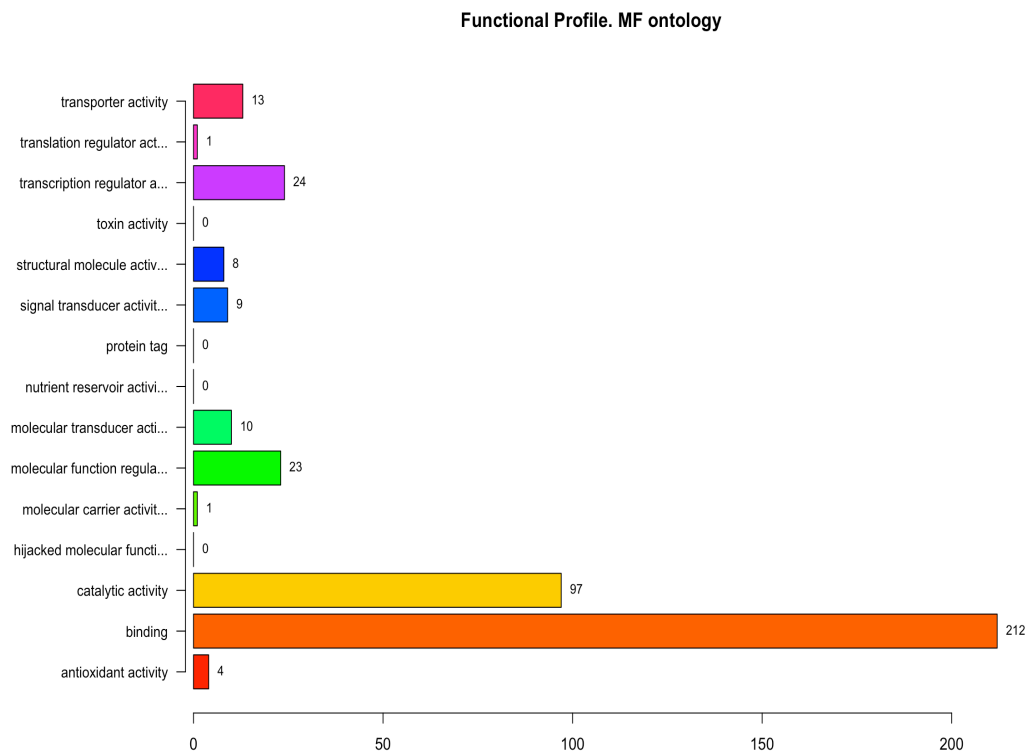


Fig. 17- Barplot de DEGs activados, GO función molecular (MF).

El análisis de GO señala que los DEGs inhibidos están enriquecidos en los componentes celulares (CC) siguientes: orgánulos, membrana, complejo macromolecular y región extracelular (Fig. 18). Los procesos biológicos (BP) más abundantes para los DEGs inhibidos son: regulación de proceso biológico (negativo y positivo), organización componente celular o biogénesis, proceso de desarrollo, respuesta a estímulo, localización, proceso metabólico y señalización entre otros (Fig. 19). En cuanto a la función molecular (MF), los DEGs inhibidos están enriquecidos principalmente en binding y actividad catalítica además de actividad reguladora de transcripción, regulación de la función molecular y actividad transportadora (Fig. 20).

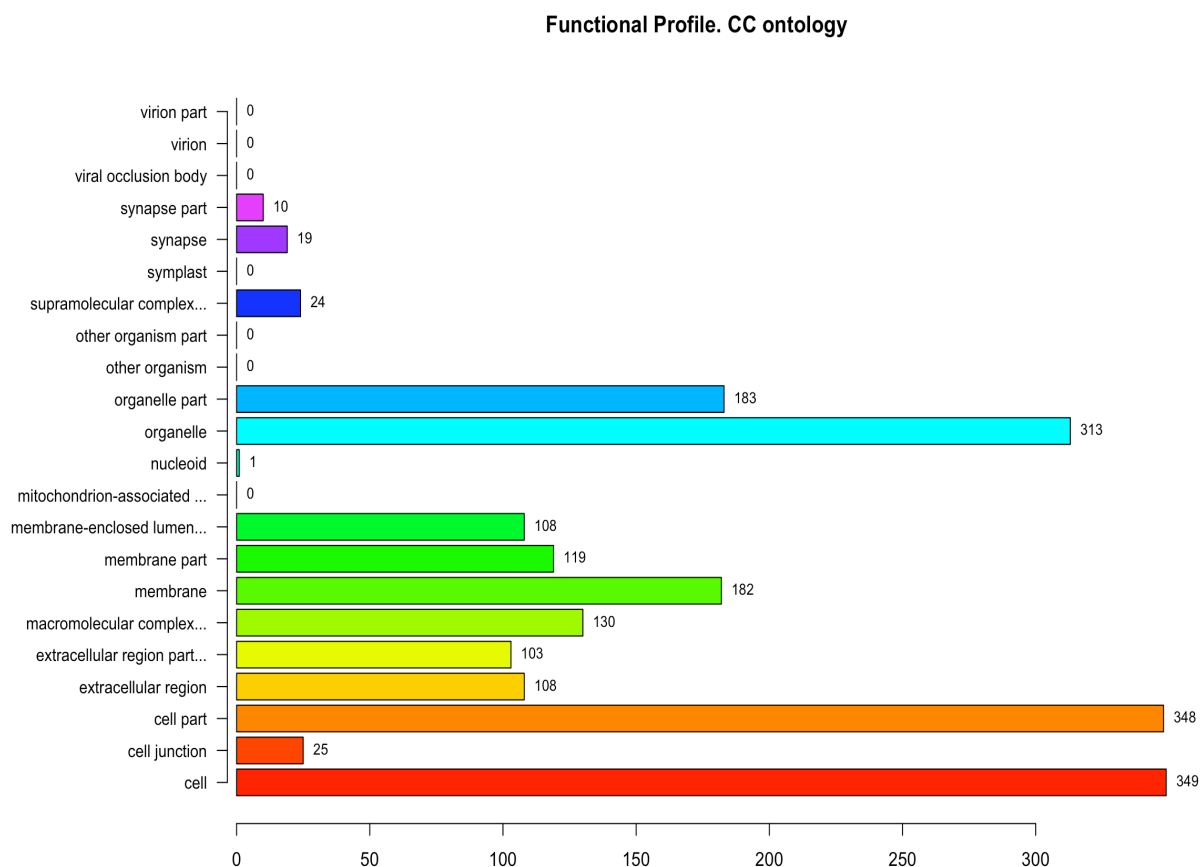


Fig. 18- Barplot de DEGs inhibidos, GO componente celular (CC).

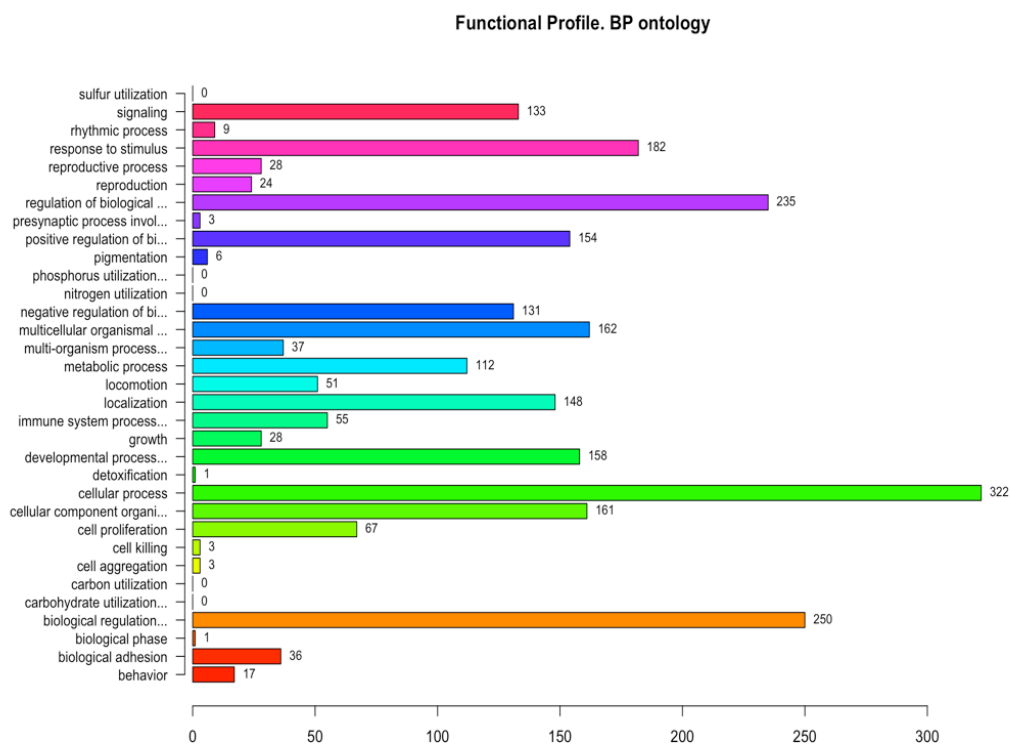


Fig. 19- Barplot de DEGs inhibidos, GO proceso biológico (BP).

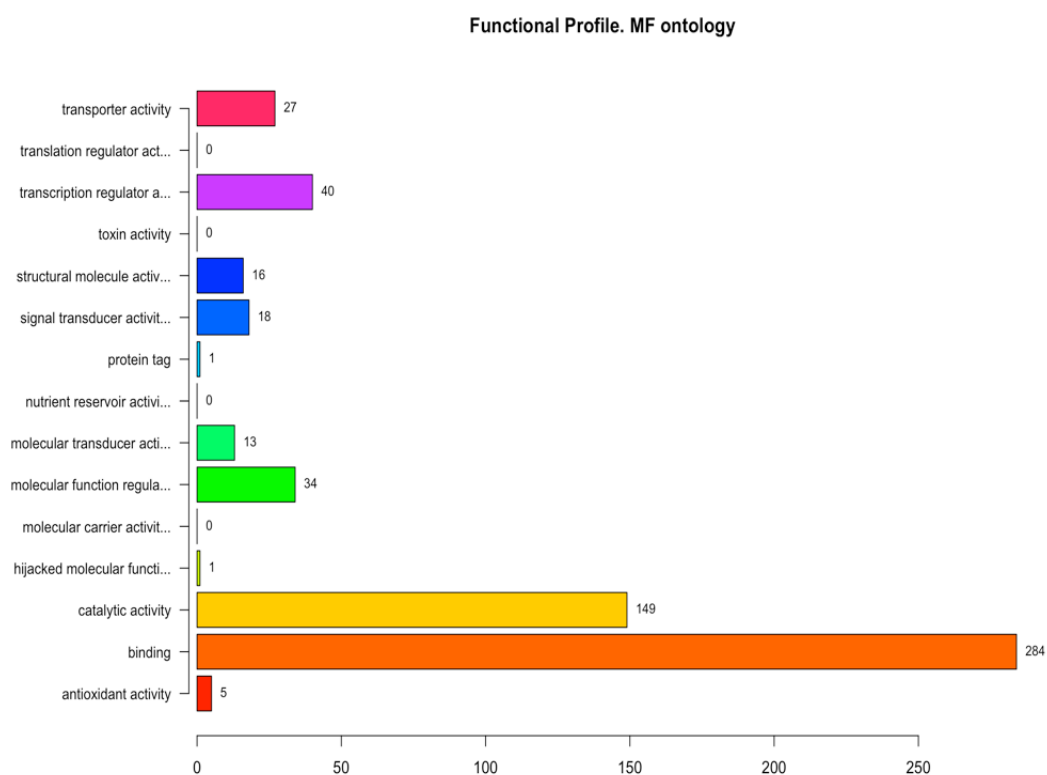


Fig. 20- Barplot de DEGs inhibidos, GO función molecular (MF).

Realizamos la comparación entre los DEGs activados e inhibidos tanto para componente celular (CC) (Fig. 21), como para proceso biológico (BP), (Fig. 22) y función molecular (MF), (Fig. 23). Obtenemos que la diferencia entre los grupos es significativa ($p < 0.001$) sólo para los procesos biológicos (BP), (Fig. 22).

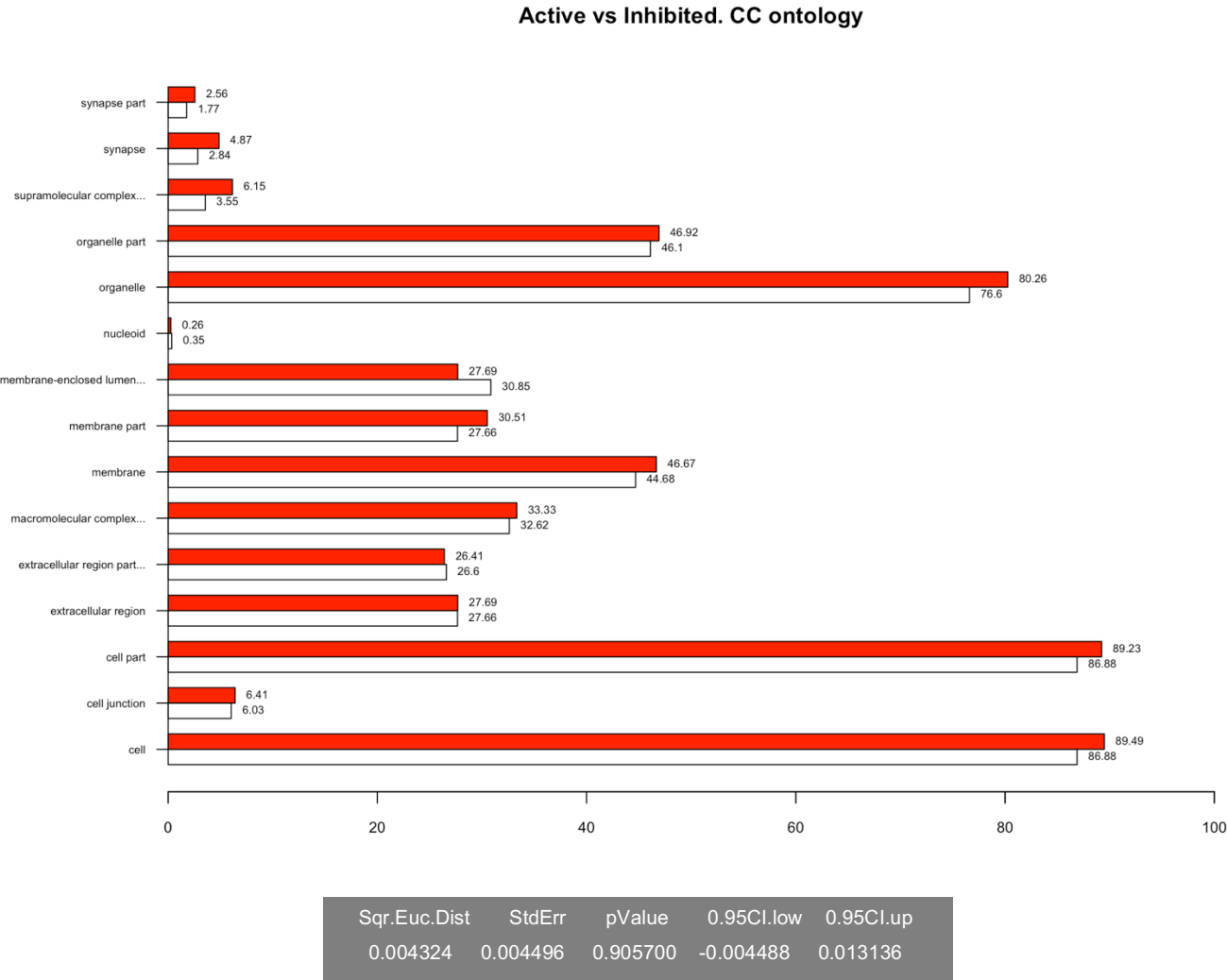
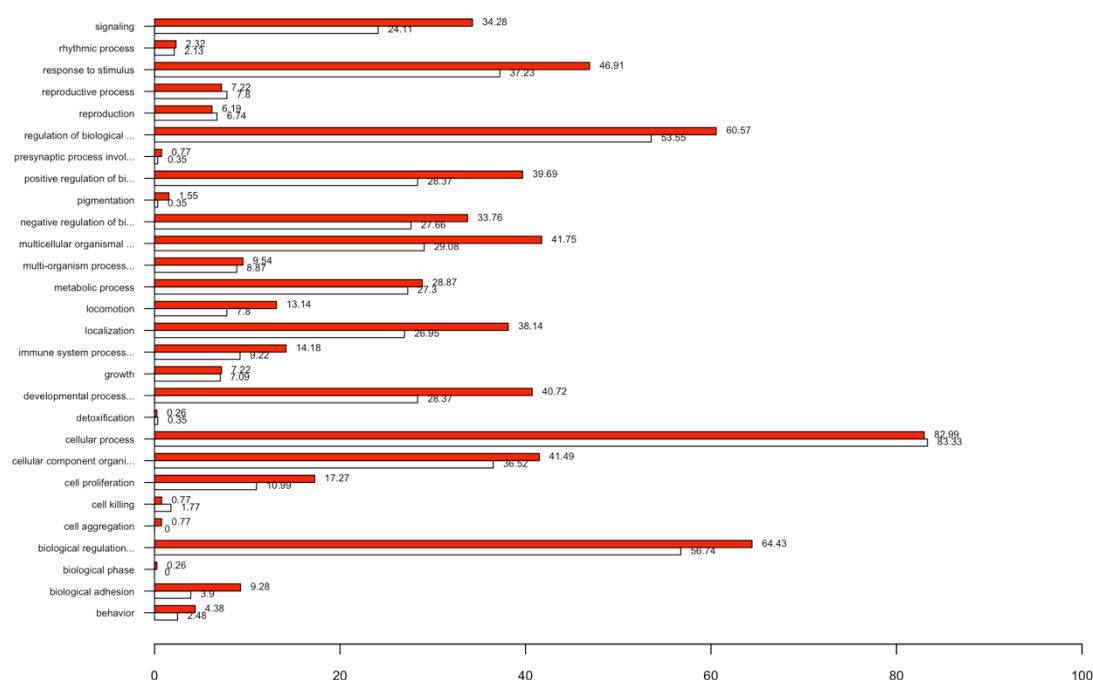


Fig. 21- Barplot comparación DEGs activados/inhibidos, GO componente celular (CC), rojo inhibidos, blanco activados.

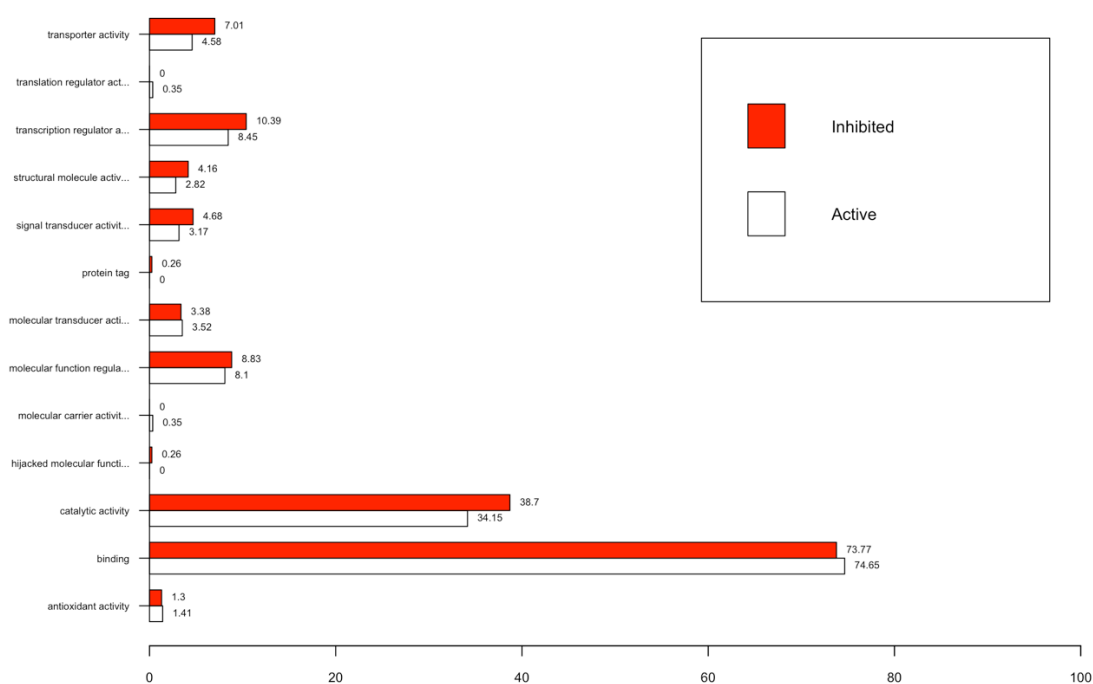
Active vs Inhibited. BP ontology



Sqr.Euc.Dist	StdErr	pValue	0.95Cl.low	0.95Cl.up
0.139059	0.059317	0.000100	0.022799	0.255318

Fig. 22- Barplot comparación DEGs activados/inhibidos, GO componente celular (BP), rojo inhibidos, blanco activados.

Active vs Inhibited. MF ontology



Sqr.Euc.Dist	StdErr	pValue	0.95CI.low	0.95CI.up
0.004725	0.004259	0.475300	-0.003622	0.013072

Fig. 23- Barplot comparación DEGs activados/inhibidos, GO componente celular (MF)

3.2.1.2- clusterProfiler

Realizamos el análisis basado en ontología de genes (GO) para el conjunto de genes expresados diferencialmente. Empleamos las funciones **groupGO**, **enrichGO**, **gseGO**.

3.2.1.2.1- groupGO

Obtenemos el perfil funcional de nuestro conjunto de genes DEGs a nivel 3 de GO con la función **groupGO** para las ontologías CC, BP y MF.

```
ggo_act_CC <- groupGO((geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "CC", level = 3, readable = FALSE)
```

El análisis de **groupGO** señala que los DEGs activados están enriquecidos en los componentes celulares (CC) siguientes: orgánulos, membrana y región extracelular (Fig. 24). Los procesos biológicos (BP) más abundantes para los DEGs activados son: procesos metabólicos, regulación procesos biológicos, procesos metabólicos compuesto de nitrógeno, regulación procesos celulares, organización procesos celulares y procesos biosintéticos entre otros (Fig. 25). En cuanto a la función molecular (MF), los DEGs activados están enriquecidos principalmente en binding de proteínas, de compuestos orgánicos cíclicos, de compuestos heterocíclicos y de iones entre otros (Fig. 26).

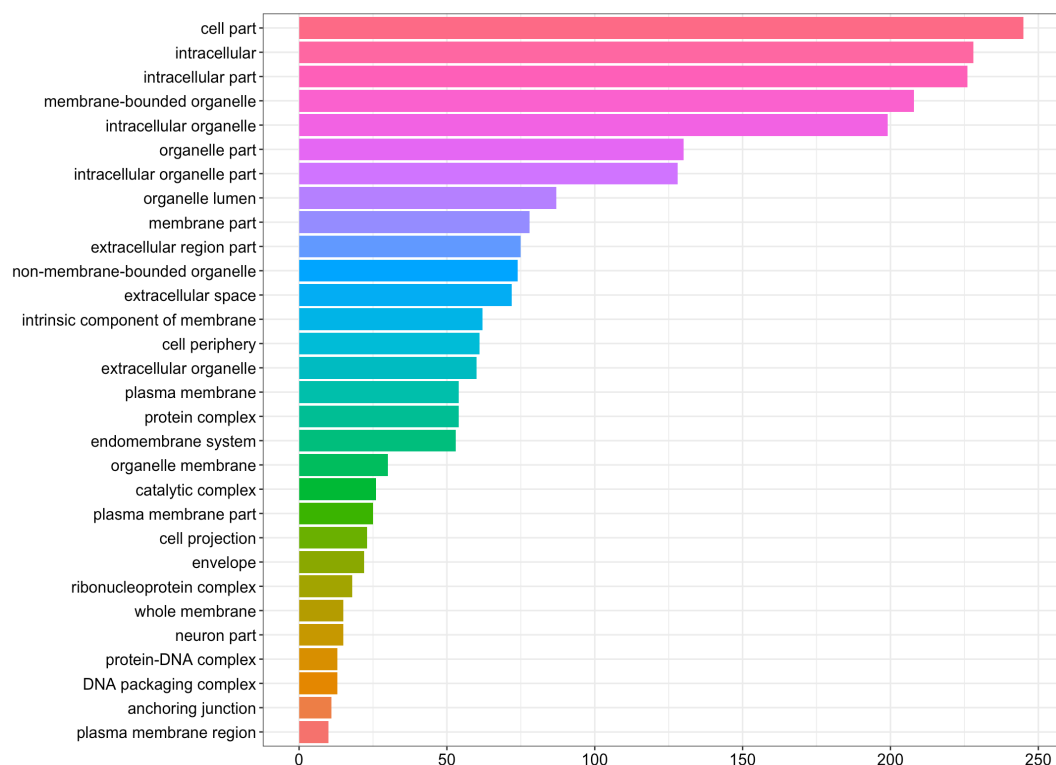


Fig. 24- Barplot de DEGs activados, groupGO componente celular (CC).

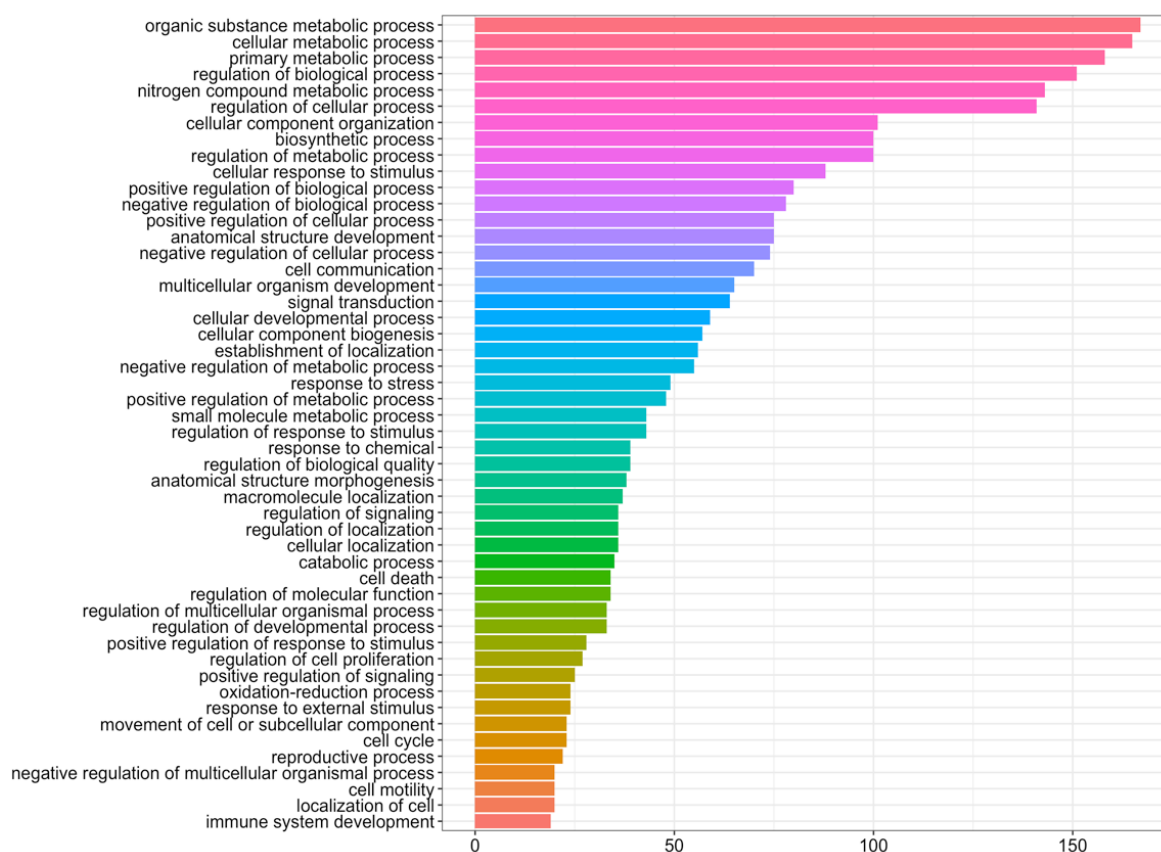


Fig. 25- Barplot de DEGs activados, groupGO procesos biológico (BP).

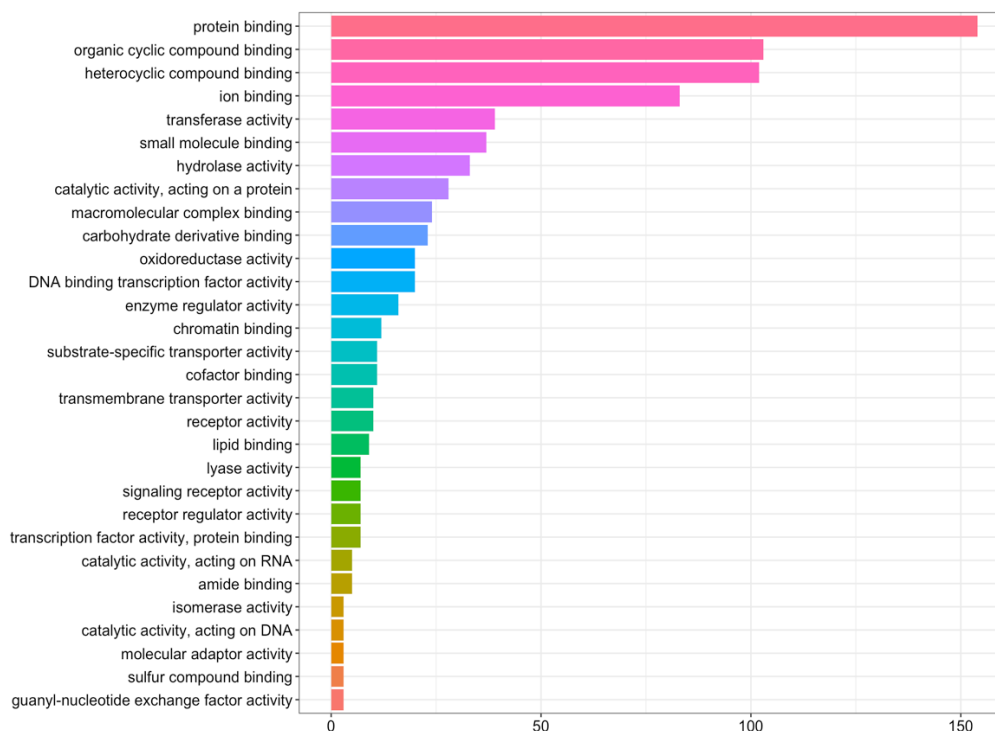


Fig. 26- Barplot de DEGs activados, groupGO función molecular (MF).

El análisis de GO señala que los DEGs inhibidos están enriquecidos en los componentes celulares (CC) siguientes: orgánulos, membrana, y región extracelular (Fig. 27). Los procesos biológicos (BP) más abundantes para los DEGs inhibidos son: procesos metabólicos, regulación procesos biológicos, procesos metabólicos compuesto de nitrógeno, regulación procesos celulares, respuesta celular a estímulo entre otros (Fig. 28). En cuanto a la función molecular (MF), los DEGs inhibidos están enriquecidos principalmente en binding de proteínas, de compuestos orgánicos cíclicos, de compuestos heterocíclicos y de iones entre otros (Fig. 29).

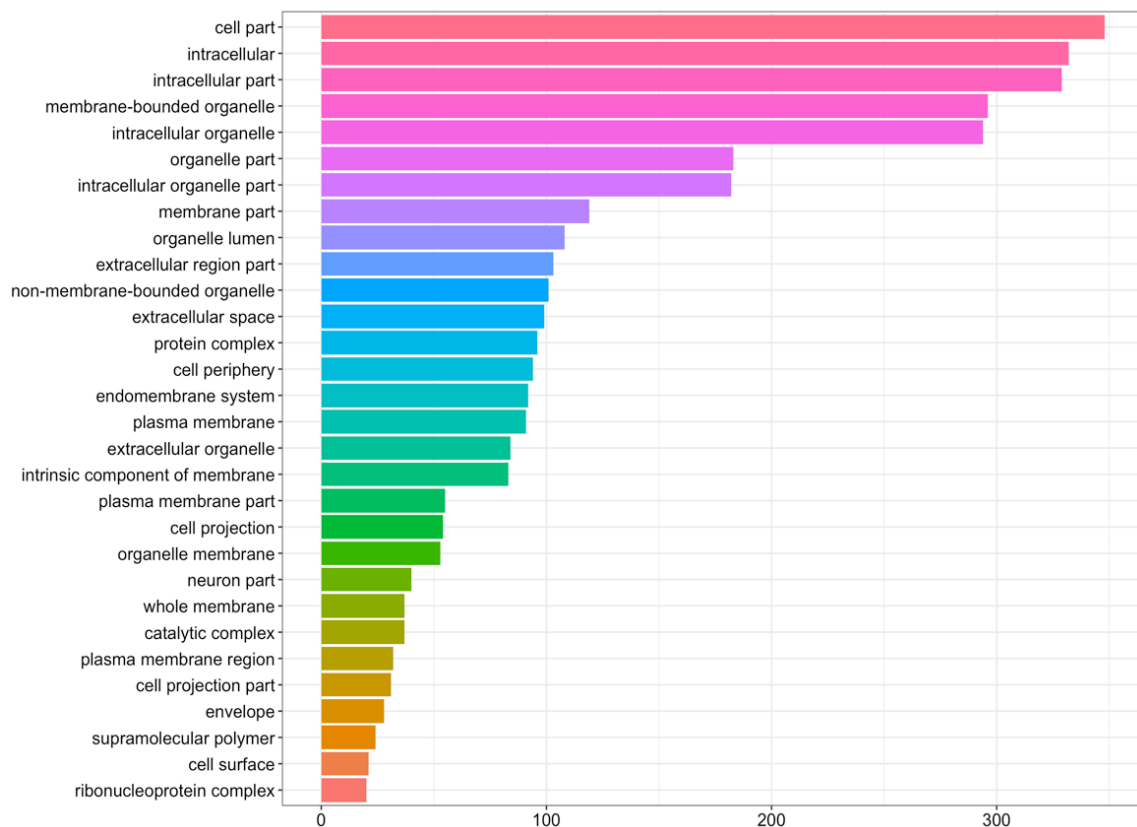


Fig. 27- Barplot de DEGs inhibidos, groupGO componente celular (CC).

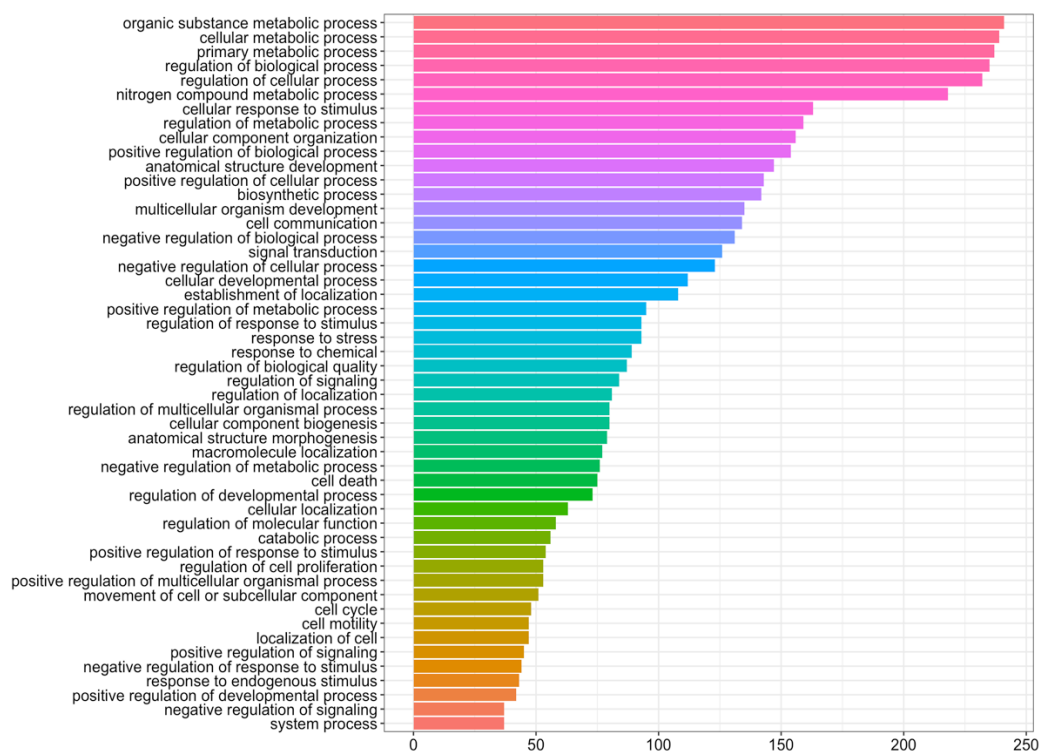


Fig. 28- Barplot de DEGs inhibidos, groupGO proceso biológico (BP).

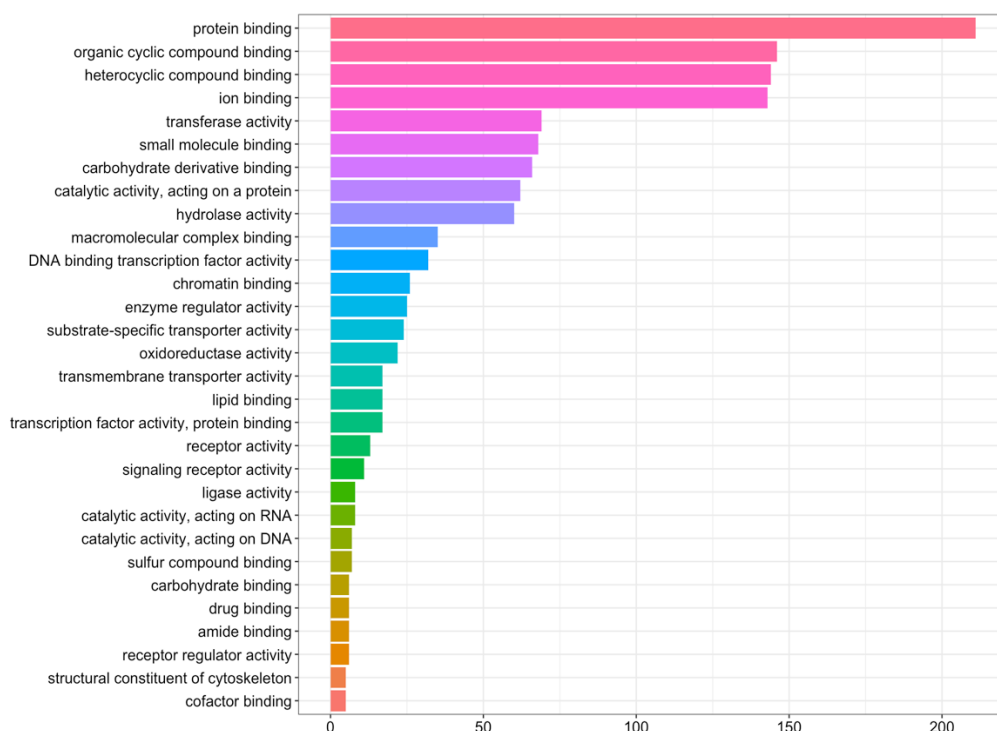


Fig. 29- Barplot de DEGs inhibidos, groupGO función molecular (MF).

3.2.1.2.2- enrichGO

Realizamos el perfil funcional de nuestro conjunto de genes DEGs de GO con la función **enrichGO** para las ontologías CC, BP y MF, esta función devuelve las categorías GO después de control FDR, pValueCutoff = 0.05.

```
ego_act_CC <- enrichGO((geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "CC", pAdjustMethod = "BH",
pvalueCutoff = 0.05, readable = FALSE)
```

El análisis de **enrichGO** señala que los DEGs activados están enriquecidos en los componentes celulares (CC) siguientes: complejo de empaquetamiento de DNA, nucleosoma, complejo proteína-DNA, cromatina nuclear, corteza celular, región citoplasmática, nucleosoma nuclear (Fig. 30). Los procesos biológicos (BP) más abundantes para los DEGs activados son: ensamblaje de cromatina, empaquetamiento DNA, ensamblaje y desensamblaje de cromatina, ensamblaje nucleosoma, cambio conformación DNA, organización nucleosoma, ensamblaje complejo DNA-proteína y organización de la subunidad del complejo proteína-DNA (Fig. 31). En cuanto a la función molecular (MF), los DEGs activados están enriquecidos principalmente en

actividad liasa carbono-sulfuro, binding de cadherina y binding de moléculas de adhesión celular (Fig. 32).

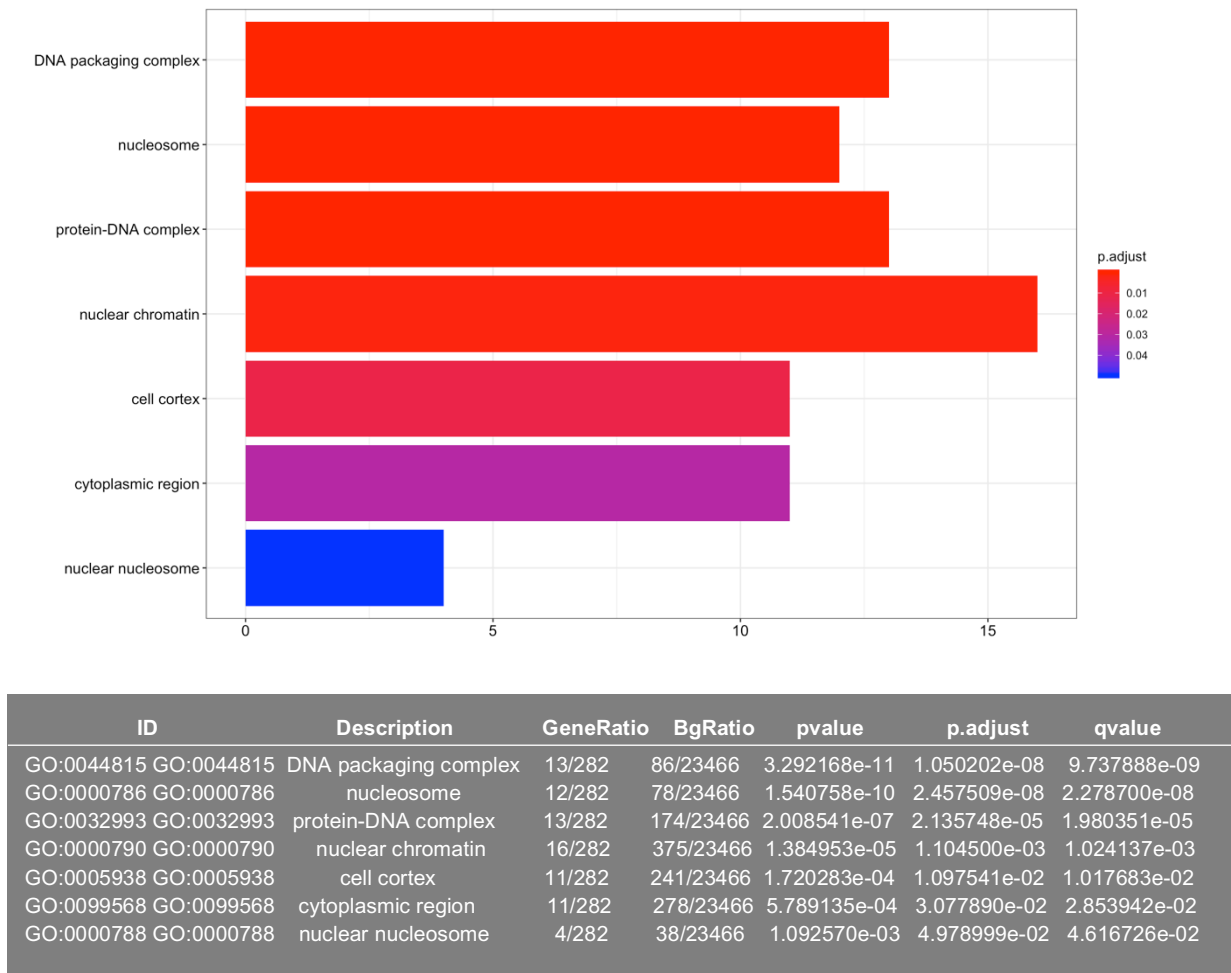
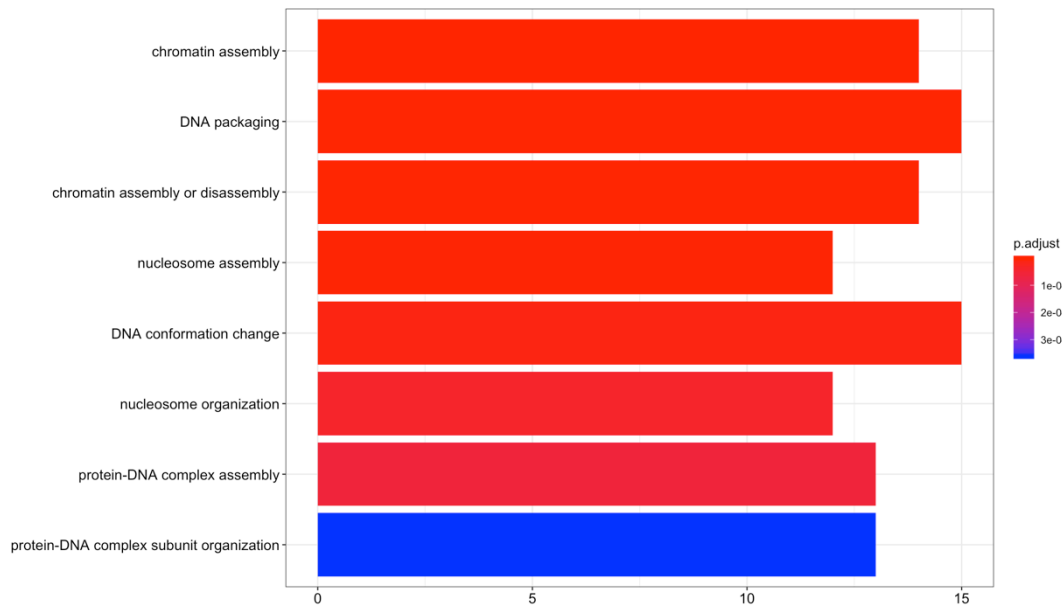
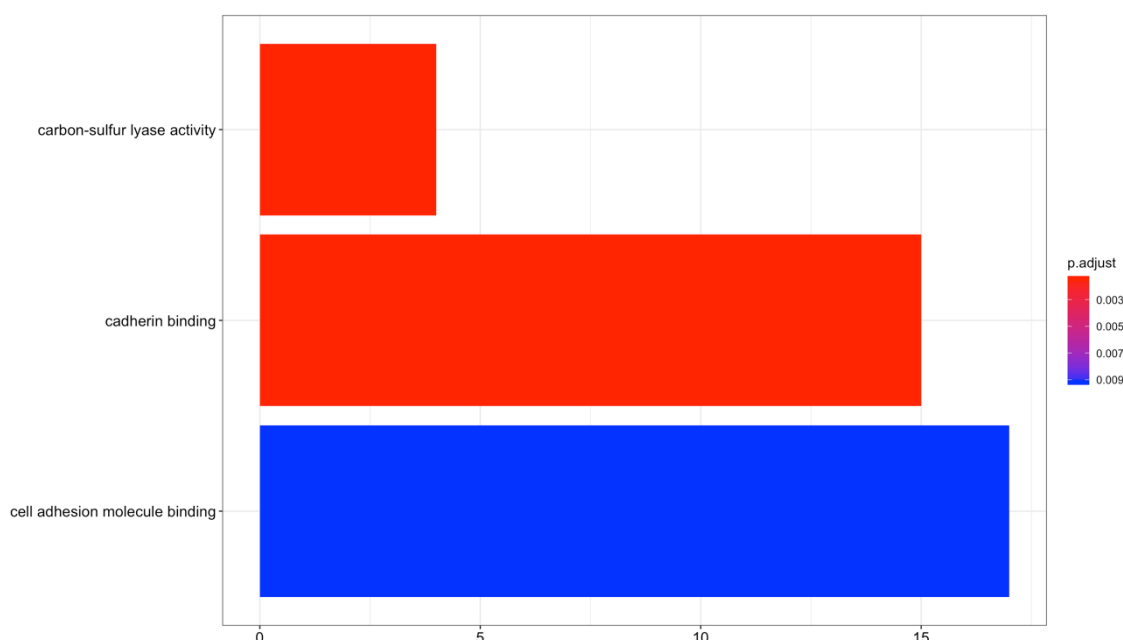


Fig. 30- Barplot y estadísticas de DEGs activados, enrihGO celular component (CC).



ID	Description	GeneRatio	BgRatio	pvalue.	p.adjust	qvalue
GO:0031497 GO:0031497	chromatin assembly	14/282	123/23577	2.675983e-10	7.366982e-07	6.954740e-07
GO:0006323 GO:0006323	DNA packaging	15/282	168/23577	1.896386e-09	2.193564e-06	2.070816e-06
GO:0006333 GO:0006333	chromatin ass/disas.	14/282	145/23577	2.390371e-09	2.193564e-06	2.070816e-06
GO:0006334 GO:0006334	nucleosome asse.	12/282	105/23577	4.929286e-09	1.127710e-05	1.064606e-05
GO:0071103 GO:0071103	DNA conf. change	15/282	200/23577	2.048148e-08	3.590634e-05	3.389709e-05
GO:0034728 GO:0034728	nucleosome organiz	12/282	134/23577	7.825573e-08	3.590634e-05	3.389709e-05
GO:0065004 GO:0065004	protein-DNA assem.	13/282	171/23577	1.555315e-07	6.116831e-05	5.774545e-05
GO:0071824 GO:0071824	protein-DNA c.su. org.	13/282	202/23577	1.050441e-06	3.614830e-04	3.412551e-04

Fig. 31- Barplot y estadísticas de DEGs activados, enriGO proceso biológico (BP).



ID	Description	GeneRatio	BgRatio	pvalue.	p.adjust	qvalue
GO:0016846 GO:0016846	carbon-sulfur lyase act.	4/284	10/22933	4.560287e-06	0.001429761	0.001351308
GO:0045296 GO:0045296	cadherin binding	15/284	302/22933	6.084088e-06	0.001429761	0.001351308
GO:0050839 GO:0050839	cell adh. mol. bind.	17/284	455/22933	5.862915e-05	0.009185234	0.008681229

Fig. 32- Barplot y estadísticas de DEGs activados, enriGO función molecular (MF).

El análisis de **enrichGO** señala que los DEGs inhibidos están enriquecidos en los componentes celulares (CC) siguientes: parte apical de la célula, membrana plasmática apical, membrana vacuolar, vacuola lítica, lisosoma y cuerpo de inclusión entre otros (Fig. 33). Los procesos biológicos (BP) más abundantes para los DEGs inhibidos son: estructura anatómica homeostasis, respuesta a la transformación del factor de crecimiento β , morfogénesis del tubo, respuesta a partícula de lipoproteína, regulación positiva motilidad celular, respuesta celular al factor de crecimiento transformante del estímulo β entre otros (Fig. 34). En cuanto a la función molecular (MF),

los DEGs activados están enriquecidos únicamente en actividad transferasa, transfiriendo grupos glicosilo. (Fig. 35).

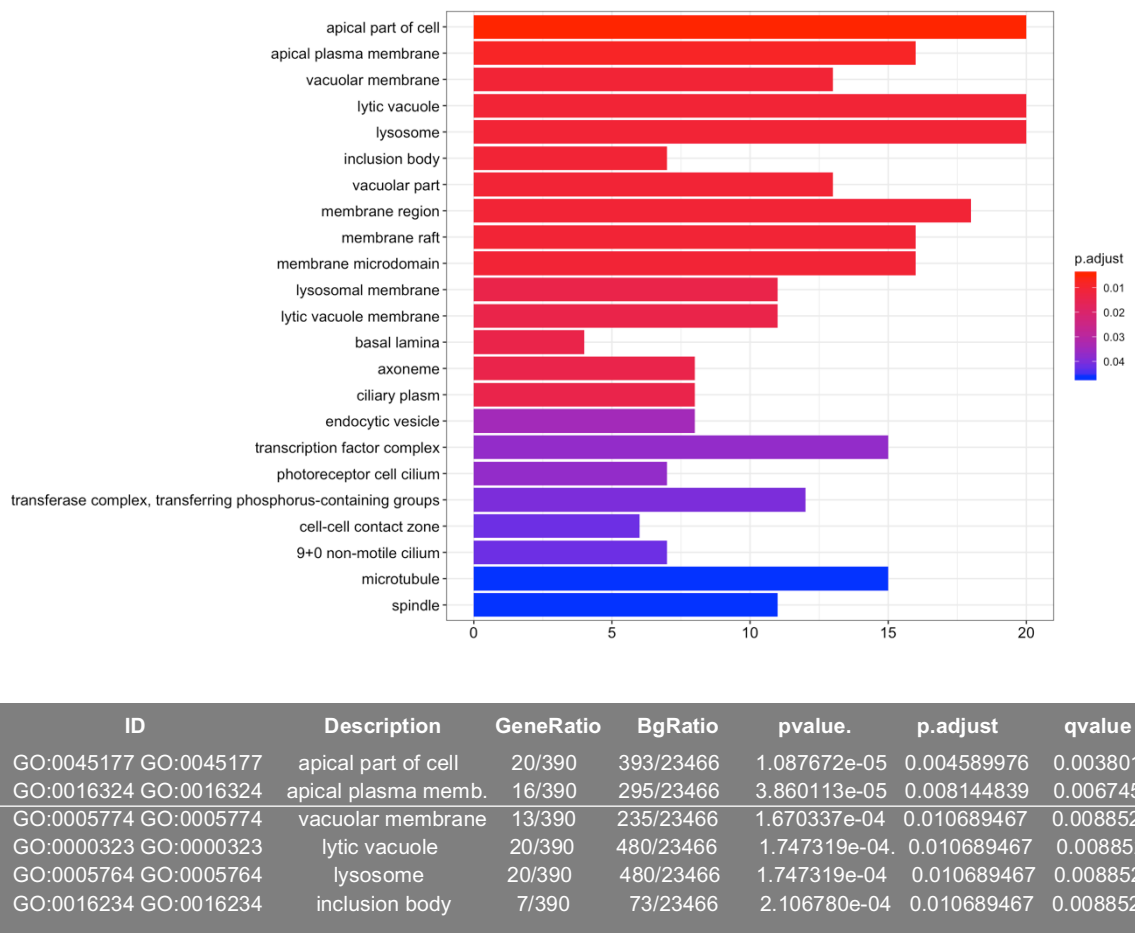


Fig. 33- Barplot y estadísticas de DEGs inhibidos, enrihGO celular component (CC).

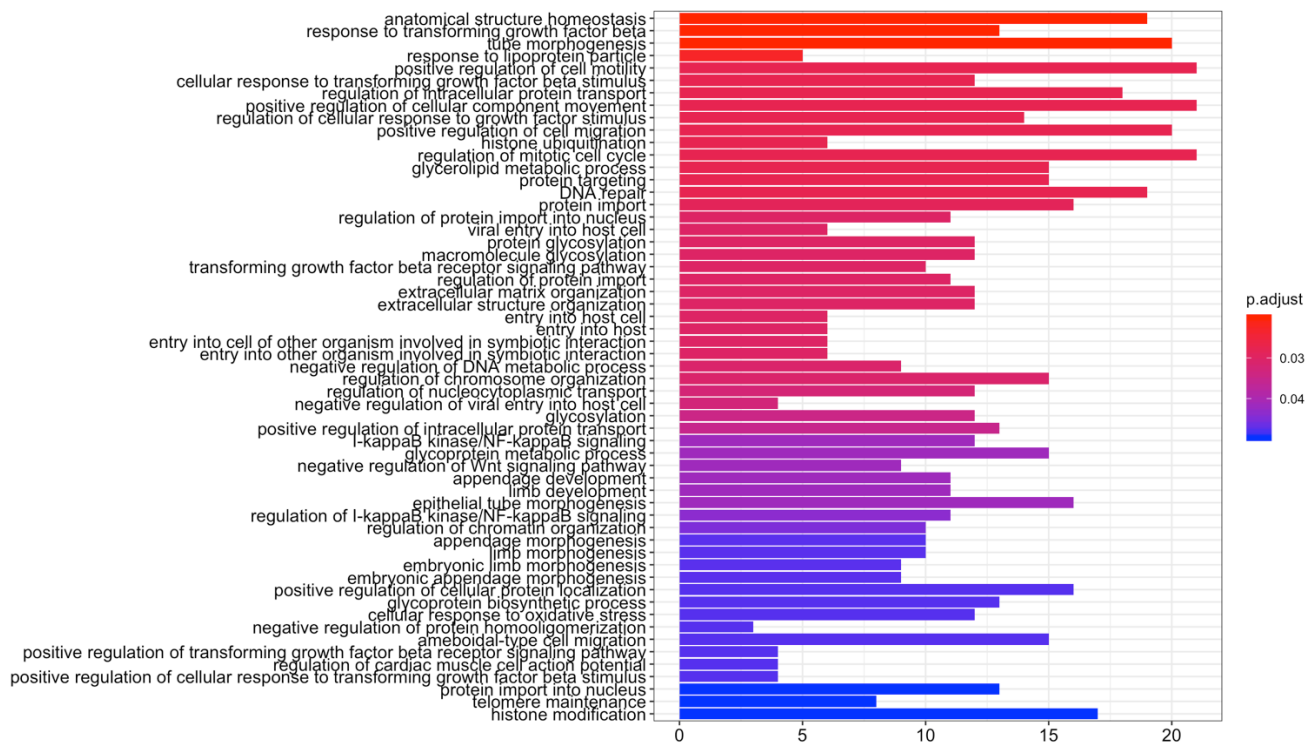


Fig. 34- Barplot y estadísticas de DEGs inhibidos, enrcihGO celular component (BP).

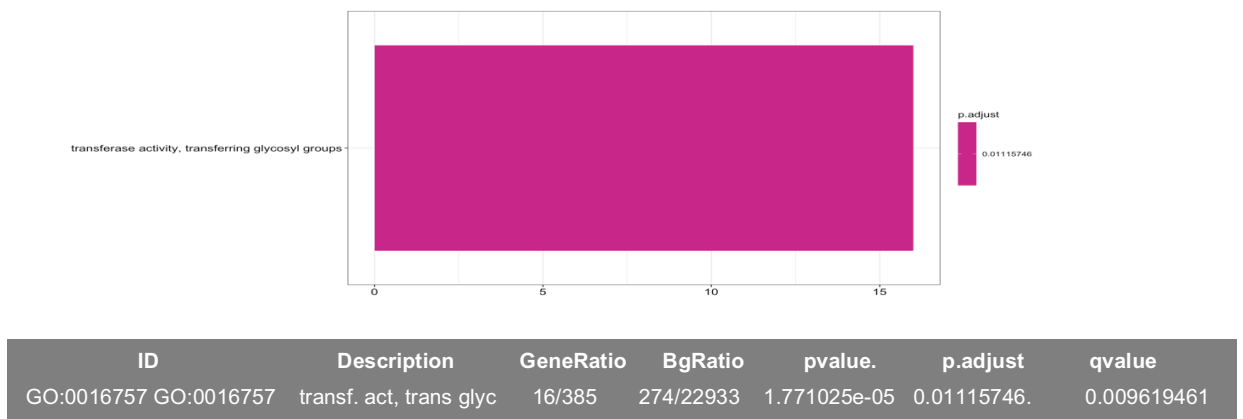


Fig. 35- Barplot y estadísticas de DEGs inhibidos, enrcihGO función molecular (MF).

Con el fin de poder saber y observar de forma más clara qué genes son los responsables de los enriquecimientos observados con anterioridad en CC, BP y MF; realizamos los **cnetplot** de los **enrichGO** obtenidos:

```
cnetplot(ego1_act_CC, foldChange = geneList_act)

cnetplot(ego1_act_CC, foldChange = geneList_act, circular = TRUE, colorEdge = TRUE)
```

La forma clásica de **cnetplot** para CC agrupa nuestros DEGs activados en 3 clusters diferenciados: 1) complejo empaquetamiento DNA, nucleosoma y complejo proteína-DNA (con la participación principalmente de histonas); 2) cromatina nuclear (con histonas y Stat3 entre otros); 3) corteza celular (Crip2, Eno1b, Pfn1 entre otros) ([Fig. 36, superior](#)). La versión circular de **cnetplot** nos permite observar en cuantos CC se encuentran los DEGs activados, así por ejemplo nos encontramos que la mayor parte de las histonas se encuentran en 4 de los 5 componentes más representados, por ejemplo, Hist1h1c e Hist1h3a entre otras ([Fig. 36, inferior](#)). Al realizar los **cnetplot** para BP obtenemos un cluster central en el cual se agrupan, ensamblaje de cromatina, empaquetamiento DNA, ensamblaje y desensamblaje de cromatina, ensamblaje nucleosoma, cambio conformación DNA, con la participación principalmente de histonas ([Fig. 37, superior](#)). El **cnetplot** circular nos indica que la mayor parte de las histonas participan en los cinco procesos biológico (BP) obtenidos en el análisis de **enrichGO** ([Fig. 37, inferior](#)). Cuando analizamos el **cnetplot** clásico para MF de los DEGs activados se diferencian 2 cluster que no tienen ningún DEG en común, totalmente separados: 1) actividad liasa carbono-sulfuro (Mgst3, Kyat3, Cenpv, Glo1); 2) binding de cadherina y binding de moléculas de adhesión celular (Eno1b, H1fx, Pfn1, Hist1h3d, Hist1h3a, Notch3 entre otros), ([Fig. 38, superior](#)).

La forma clásica de **cnetplot** para CC agrupa nuestros DEGs inhibidos en 2 clusters diferenciados: 1) parte apical de la célula, membrana plasmática apical (con la participación de Slc12a2, Fat1, Myo6, Shroom2 entre otros); 2) membrana vacuolar, vacuola lítica, lisosoma (con Vps35, Ahnak, Lamp2 entre otros); ([Fig. 39, superior](#)). La versión circular de **cnetplot** nos muestra que los DEGs inhibidos que se encuentran en mayor número de CC son, Myo7a, Hsp90ab1, Acpp, Lrp2 entre otros); ([Fig. 39, inferior](#)). Al realizar los **cnetplot** para BP obtenemos 5 clusters, uno por cada BP: 1) respuesta a partícula de lipoproteína (Npc1, Lpl), 2) morfogénesis del tubo (Stk4, Slc12a2), 3) respuesta a la transformación del factor de crecimiento β (Sox9, Hsp90ab1), 4) regulación positiva motilidad celular (Sox9, Gant2), 5) estructura anatómica homeostasis (Ctgf, Hsp90ab1), ([Fig. 40, superior](#)). El **cnetplot** muestra que casi todos los DEGs tienen un solo BP, aunque hay algunos que se encuentran en varios Sox9, Gcnt, Ctsh

entre otros (Fig. 40, inferior). Cuando analizamos el **cnetplot** clásico para MF de los DEGs inhibidos encontramos un único cluster, actividad transferasa, transfiriendo grupos glicosilo (Gcnt2, Ugcg, Alg8, Ext12, Tnks2 entre otros) (Fig. 41, superior). En este caso como sólo tenemos una categoría para MF el **cnetplot** clásico y circular son prácticamente iguales (Fig. 41, inferior).

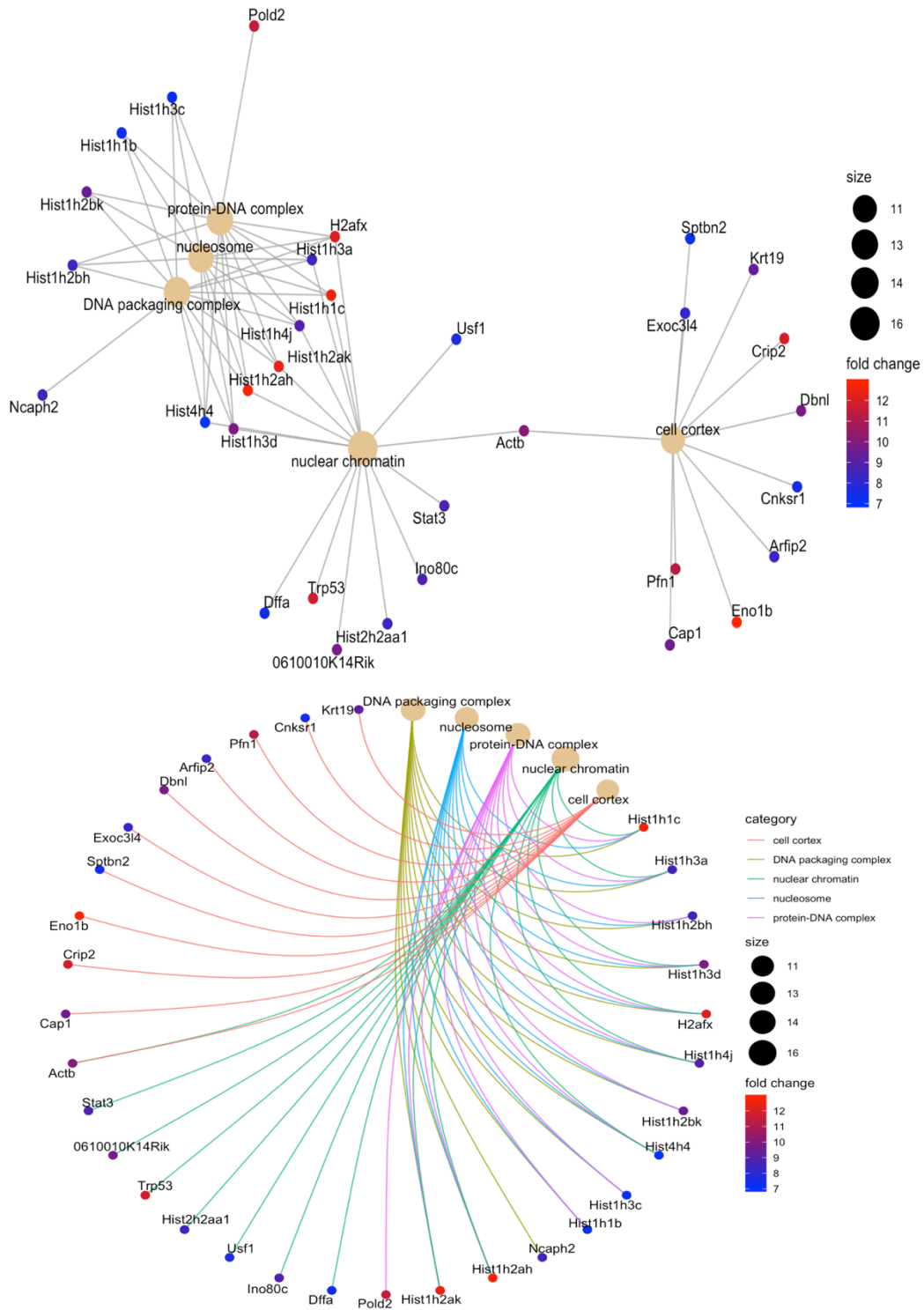


Fig. 36- cnetplot de enrchiGO DEGs activados, componente celular (CC)

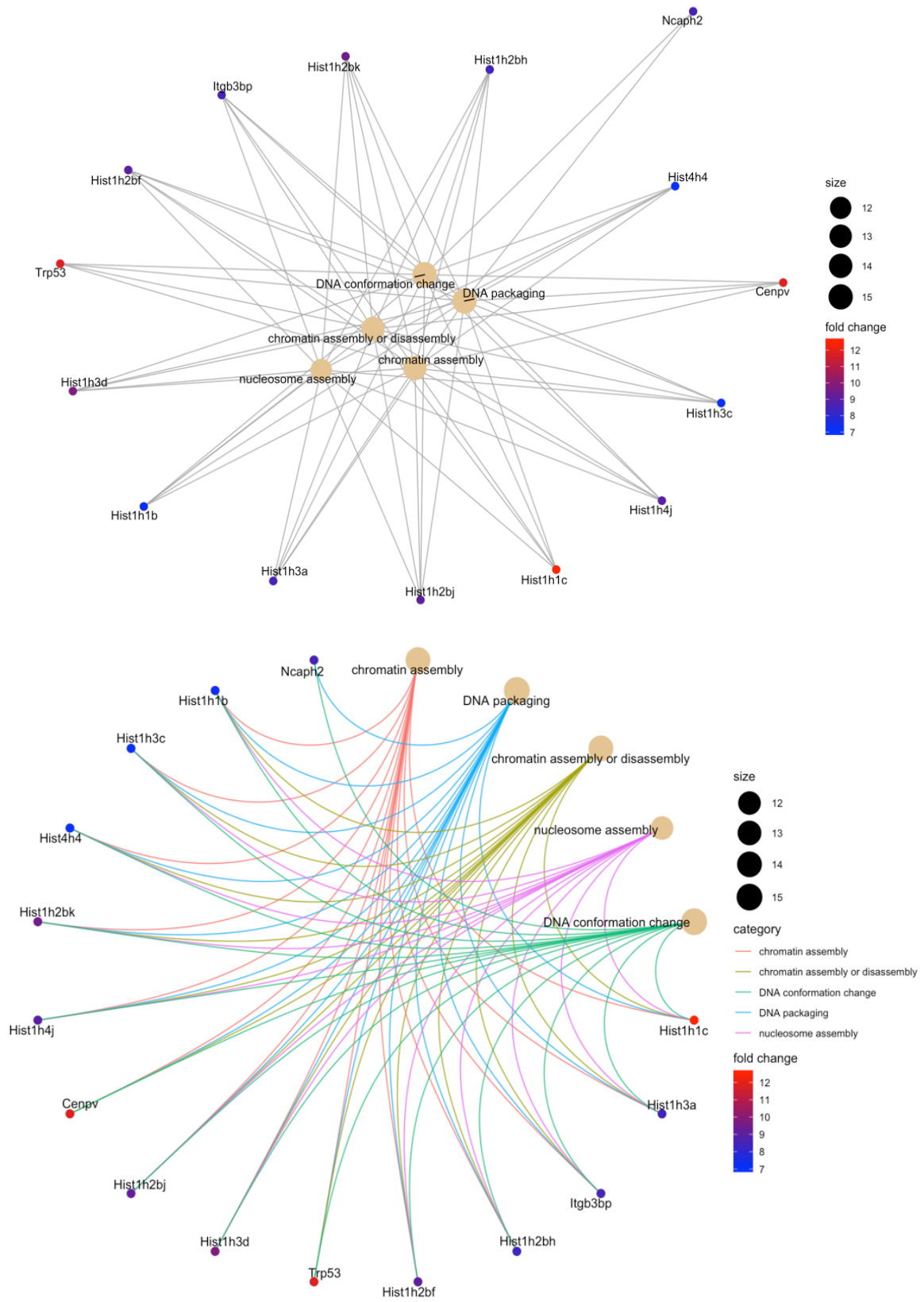


Fig. 37- cnetplot de enrchiGO DEGs activados, proceso biológico (BP)

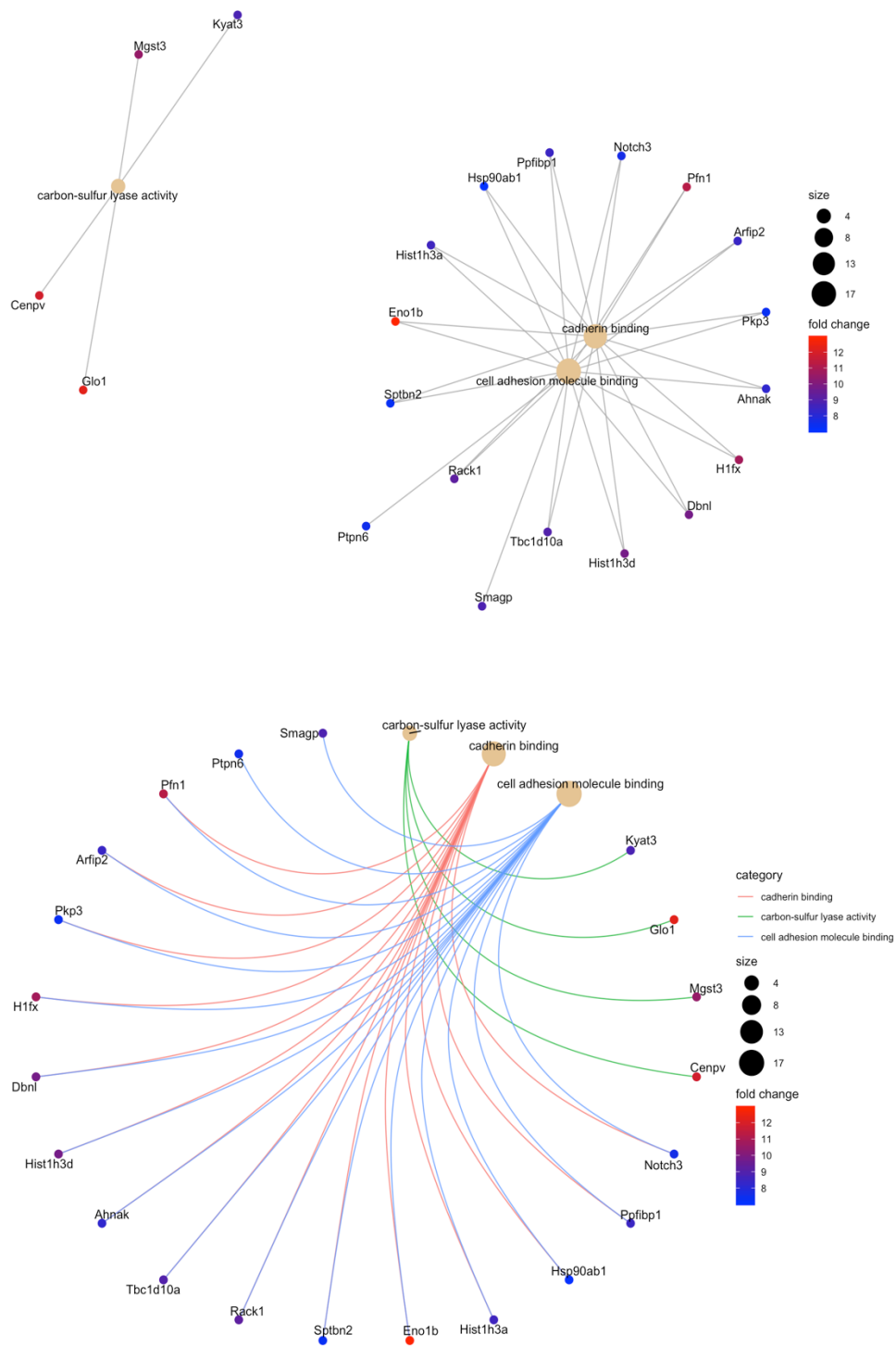


Fig. 38- cnetplot de enrchiGO DEGs activados, función molecular (MF)

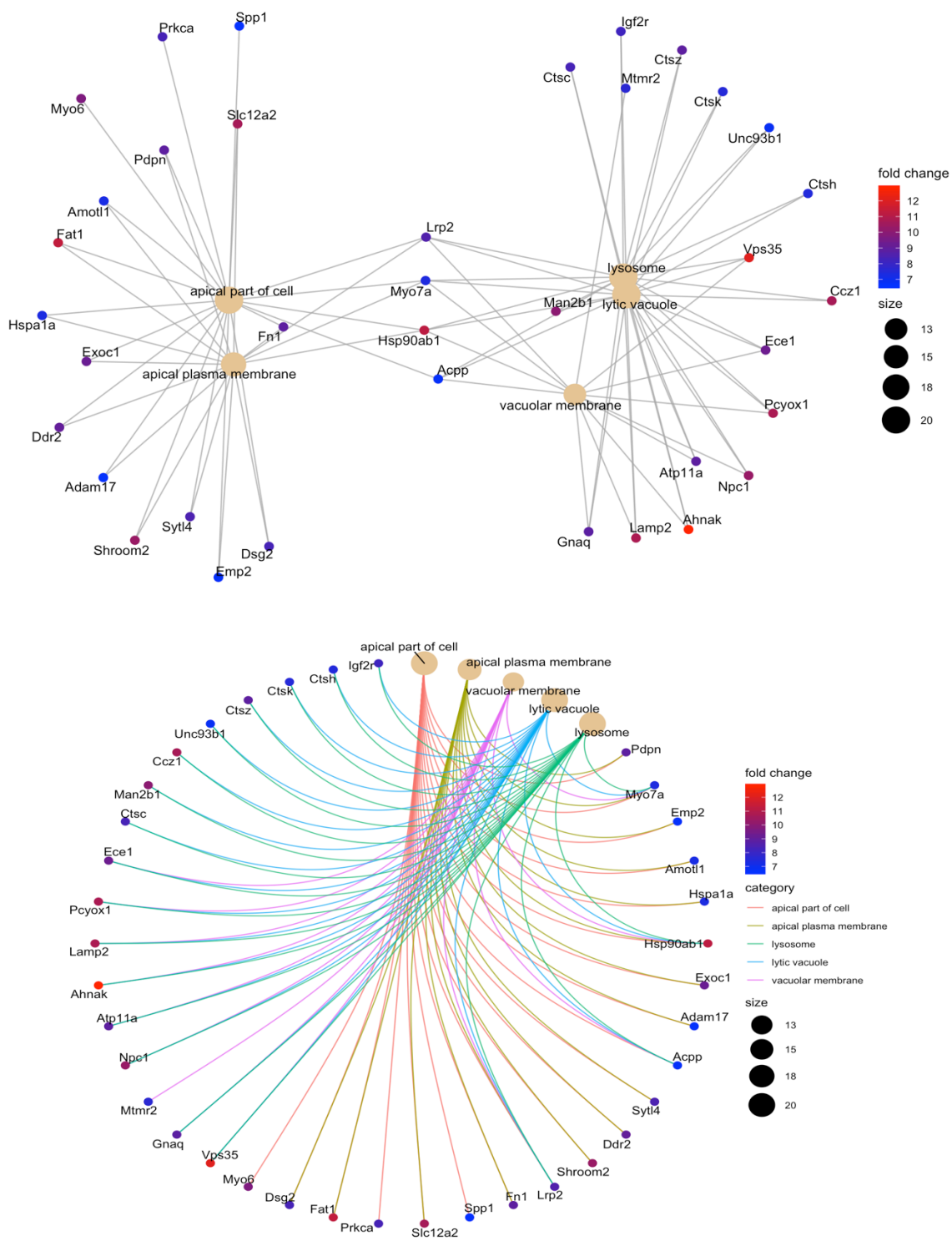


Fig. 39- cnetplot de enrichGO DEGs inhibidos, componente celular (CC)

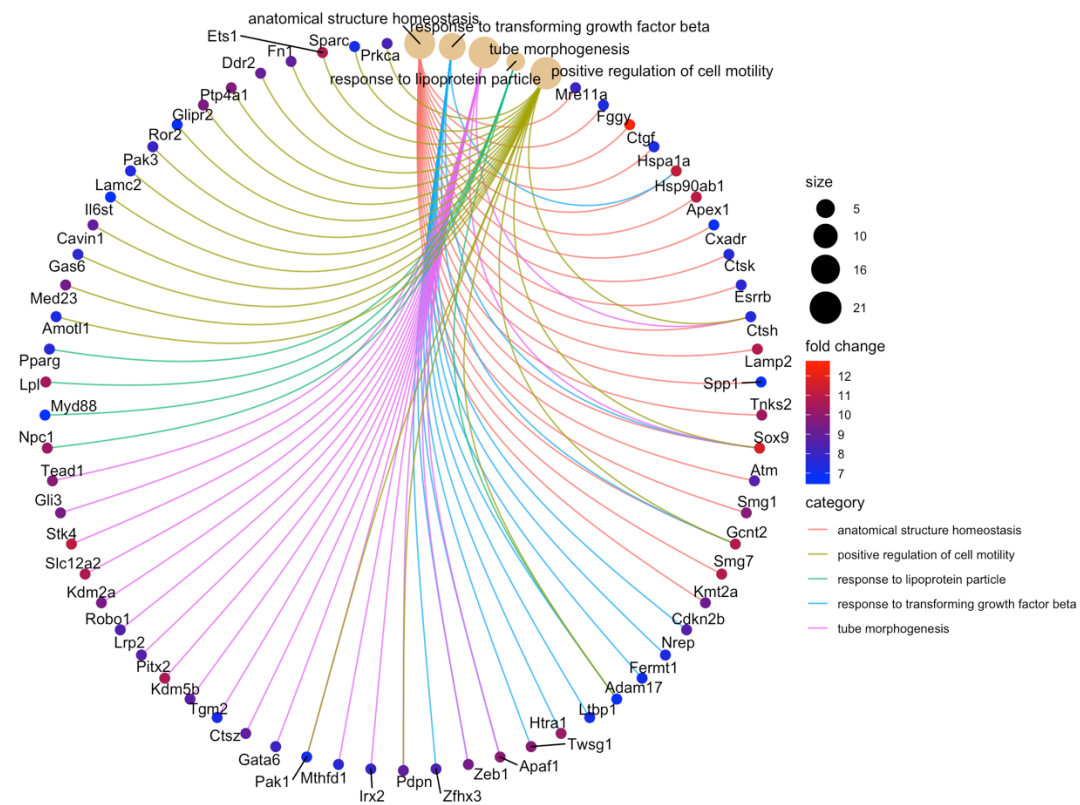
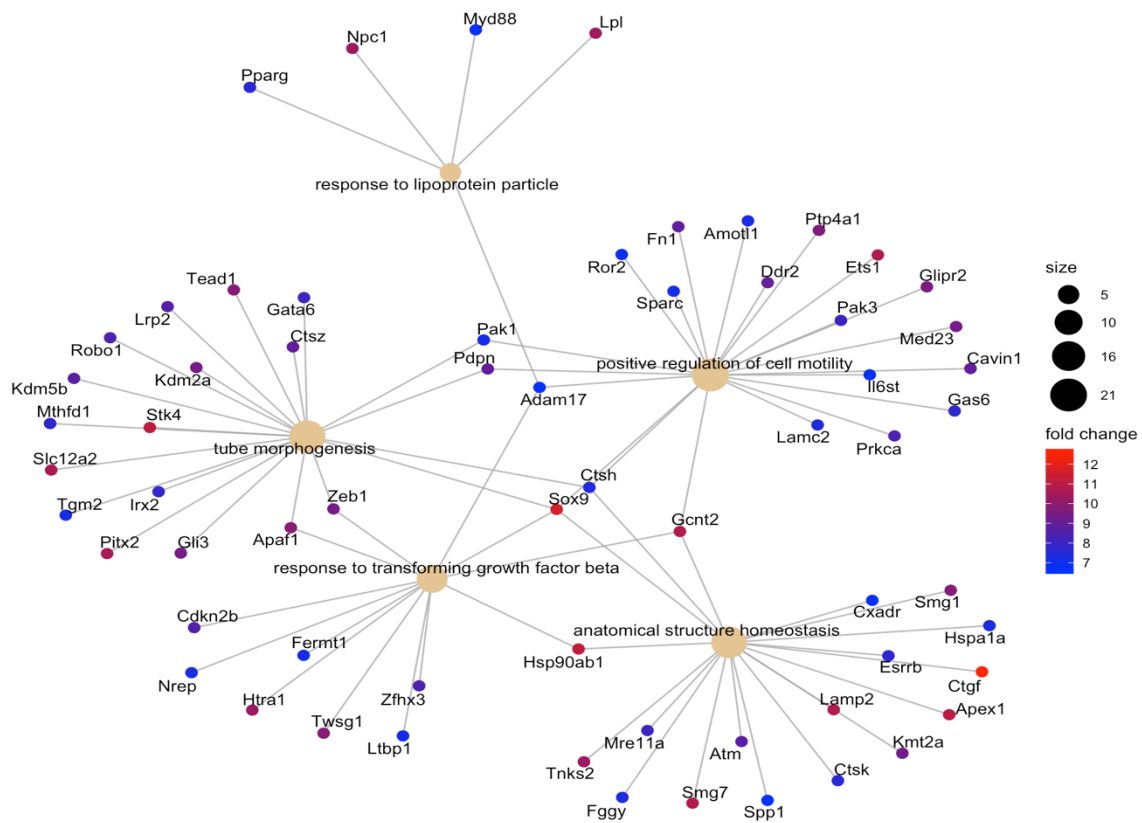


Fig. 40- cnetplot de encihGO DEGs inhibidos, proceso biológico (BP)

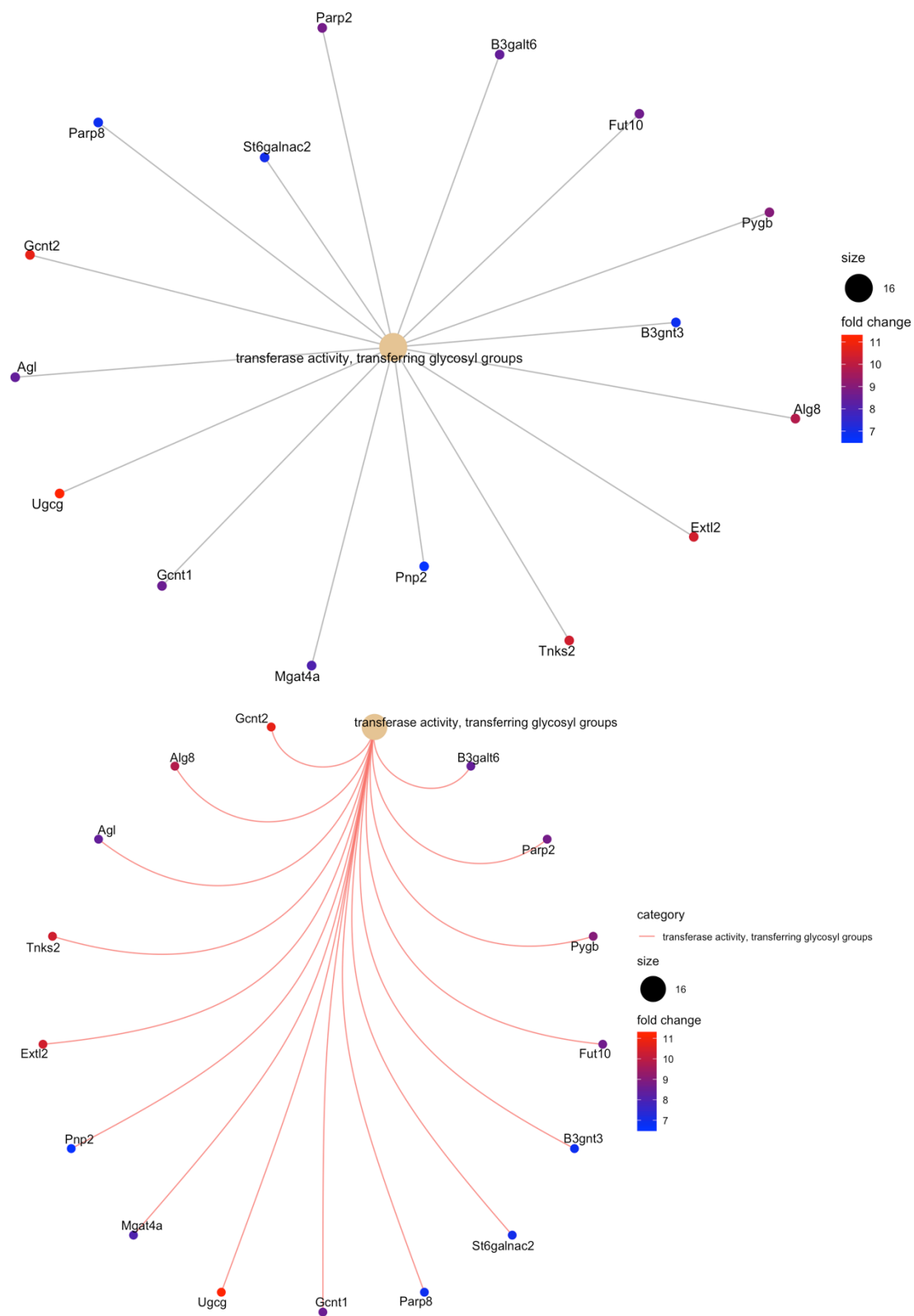


Fig. 41- cnetplot de enrichGO DEGs inhibidos, función molecular (MF)

Para poder visualizar la relación existente entre los términos de GO enriquecidos en nuestras muestras tanto a nivel de CC como BP y MF realizamos **plotGOgraph**.

El **plotGOgraph** señala que los GO para CC de los DEGs activados están interrelacionados de forma directa: 1) región citoplasmática y corteza celular y por otro lado 2) complejo macromolecular - empaquetamiento de DNA - complejo DNA-proteína – nucleosoma - nucleosoma nuclear, (Fig. 42). Cuando analizamos el **plotGOgraph** para BP de los DEGs activados vemos que existen 4 grupos de interrelación de los GO y que estos se relacionan finalmente con el ensamblaje nucleosoma. 1) cambio conformación DNA - empaquetamiento DNA, 2) ensamblaje de cromatina - ensamblaje y desensamblaje de cromatina 3) ensamblaje complejo DNA-proteína - organización nucleosoma , 4) ensamblaje complejo DNA-proteína, (Fig. 43). En cuanto a la función molecular (MF), al realizar **plotGOgraph** vemos que existen 2 grupos sin conexión entre ellos: 1) actividad liasa carbono-sulfuro y 2) binding de cadherina - binding de moléculas de adhesión celular, (Fig. 44).

El **plotGOgraph** señala que los GO para CC de los DEGs inhibidos están interrelacionados de forma directa: 1) parte apical de la célula - membrana plasmática apical, 2) complejo macromolecular – empaquetamiento, 2) región de la membrana – microdominio de membrana – balsa de membrana, 3) cuerpo de inclusión, 4) vacuola lítica – lisosoma, 5) parte de vacuola – membrana vacuolar, , (Fig. 45). Cuando analizamos el **plotGOgraph** para BP de los DEGs inhibidos vemos que existen 6 grupos de interrelación de los GO: 1) estructura anatómica homeostasis, 2) morfogénesis del tubo, 3) regulación positiva motilidad celular , 4) regulación intracelular, 5) regulación celular respuesta estímulo factor de crecimiento, 6) respuesta a la transformación del factor de crecimiento β - respuesta a partícula de lipoproteína - respuesta celular al factor de crecimiento transformante del estímulo β (Fig. 46). En cuanto a la función molecular (MF), los DEGs activados están enriquecidos únicamente en actividad transferasa, transfiriendo grupos glicosilo. (Fig. 47).

```

groupGOTerms:      GOBPTerm, GOMFTerm, GOCCTerm environments built.

Building most specific GOs .....
( 470 GO terms found. )

Build GO DAG topology .....
( 470 GO terms and 865 relations. )

Annotating nodes .....
( 23466 genes annotated to the GO terms. )

$dag
A graphNEL graph with directed edges
Number of Nodes = 35
Number of Edges = 61

```

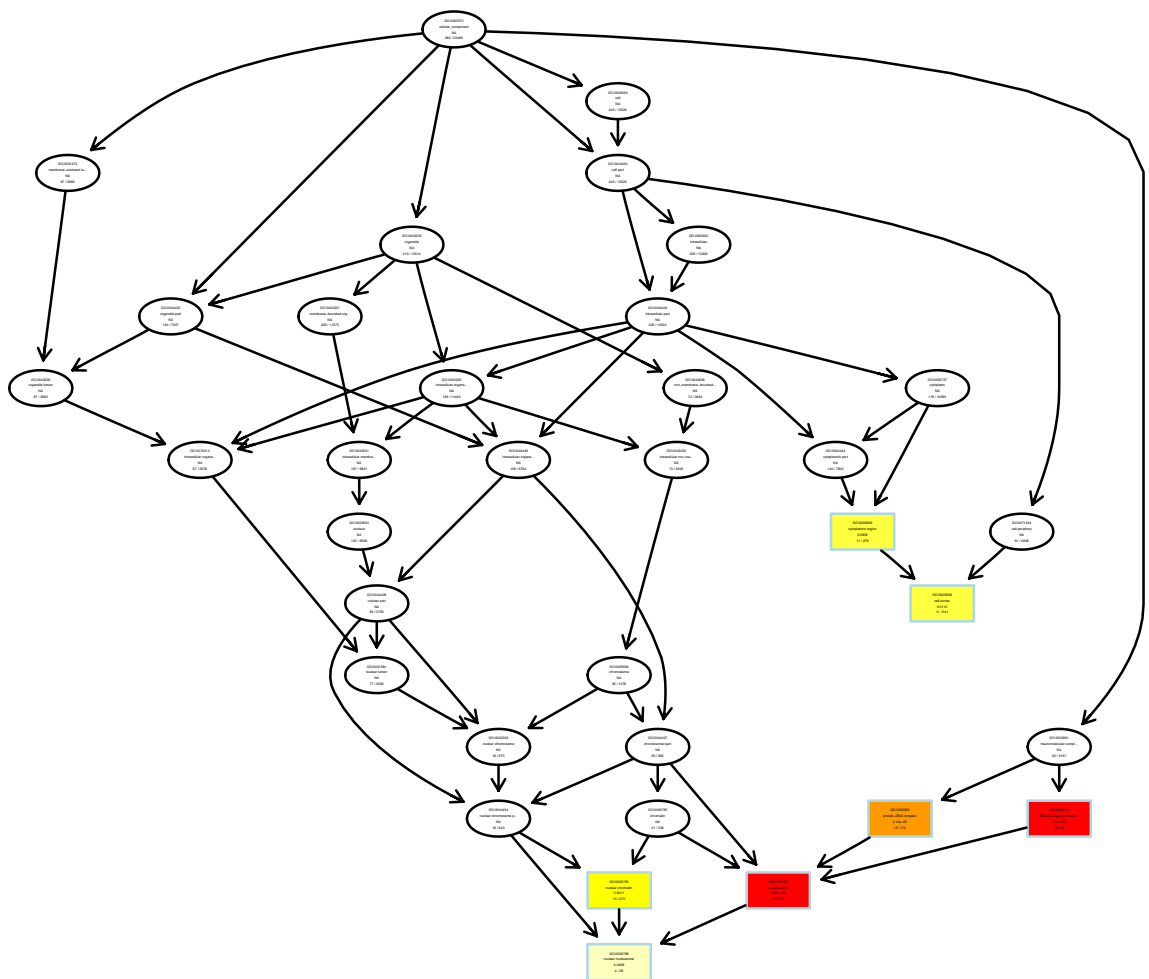


Fig. 42- plotGOgraph de enrihGO DEGs activados, componente celular (CC


```

groupGOTerms:      GOBPTerm, GOMFTerm, GOCCTerm environments built.

Building most specific GOs .....
( 719 GO terms found. )

Build GO DAG topology .....
( 719 GO terms and 920 relations. )

Annotating nodes .....
( 22933 genes annotated to the GO terms. )

$dag
A graphNEL graph with directed edges
Number of Nodes = 8
Number of Edges = 7

```

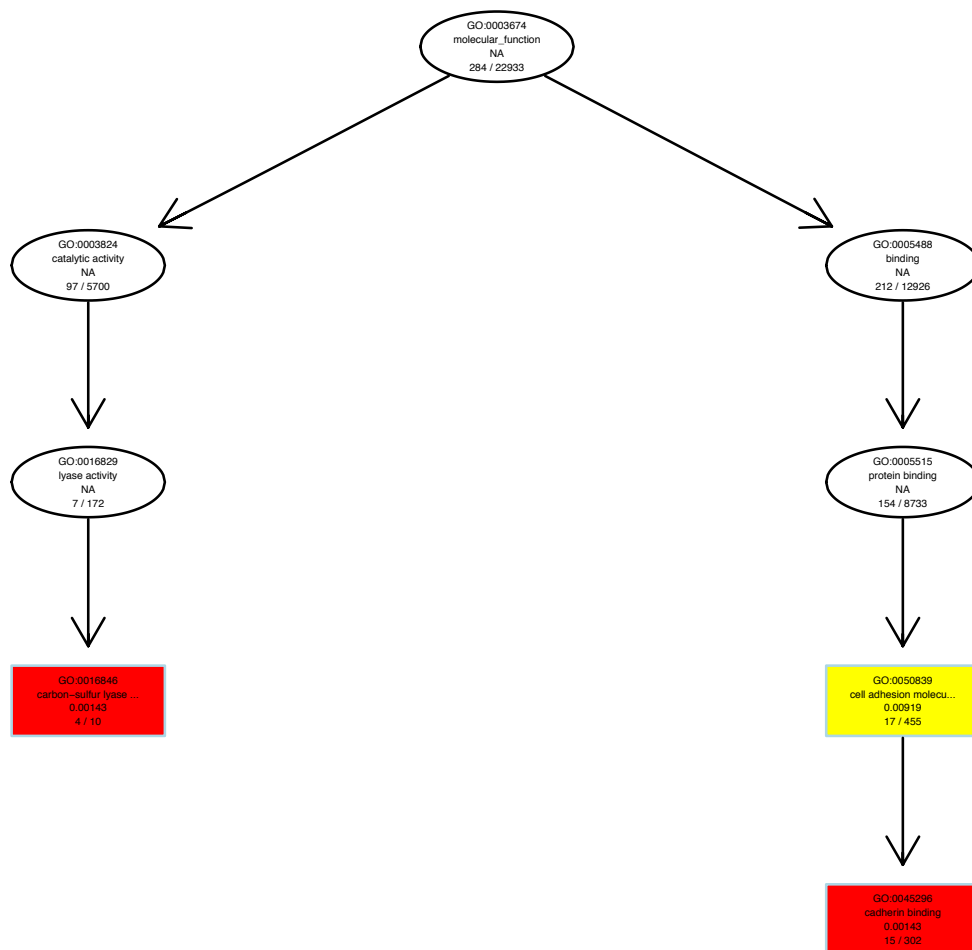


Fig. 44- plotGOgraph de encihGO DEGs activados, función molecular (MF)

groupGOTerms: GOBPTerm, GOMFTerm, GOCCTerm environments built.

Building most specific GOs
(4930 GO terms found.)

Build GO DAG topology
(4930 GO terms and 10972 relations.)

Annotating nodes
(23577 genes annotated to the GO terms.)

\$dag

A graphNEL graph with directed edges

Number of Nodes = 74

Number of Edges = 136

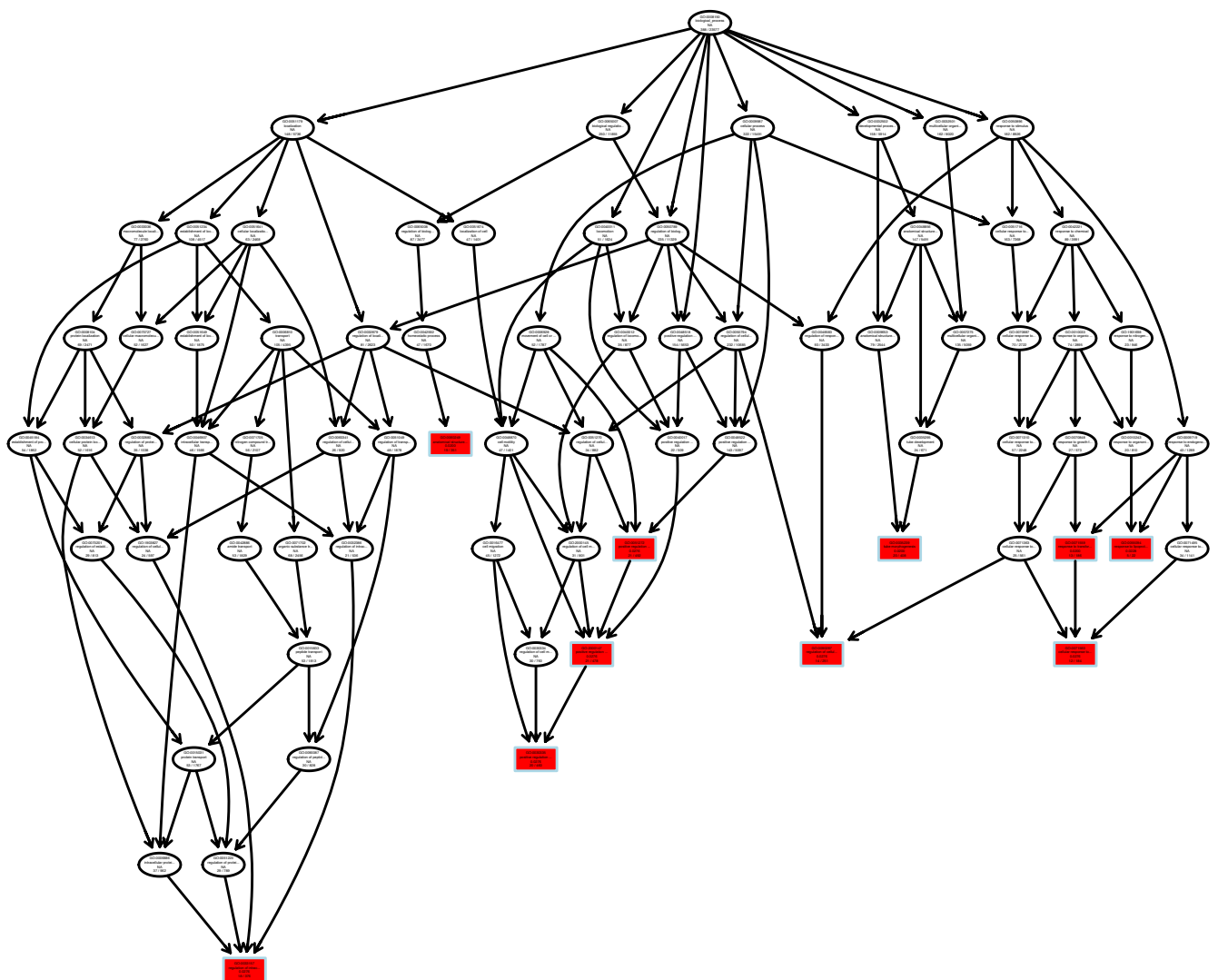


Fig. 45- plotGOgraph de encrihGO DEGs proceso biológico (BP)

```

groupGOTerms:      GOBPterm, GOMFterm, GOCCTerm environments built.

Building most specific GOs .....
( 959 GO terms found.)

Build GO DAG topology .....
(959 GO terms and 1252 relations.. )

Annotating nodes .....
( 22933 genes annotated to the GO terms.

$dag
A graphNEL graph with directed edges
Number of Nodes = 4
Number of Edges = 3

```



Fig. 46- plotGOgraph de enriehGO DEGs función molecular (MF)

3.2.1.2.2- gseGO

Realizamos el análisis de enriquecimiento de conjuntos de genes de ontología de los DEGs activados e inhibidos con la función **gseGO** para las tres ontologías, CC, BP y MF, con un pvalueCutoff = 0.05.

```

gsecc_act_CC <- gseGO(geneList_act2, ont = "CC", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
pAdjustMethod = "BH")

```

No obtuvimos en ningún caso términos gseGO, tanto para los DEGs activados como inhibidos.

3.2.2- KEGG análisis de los genes expresados diferencialmente

Utilizamos el paquete de bioconductor clusterProfiler para identificar las categorías KEGG representadas. Realizamos el análisis de los genes expresados diferencialmente, activados e inhibidos de forma independiente.

3.2.2.1- enrichKEGG

Realizamos el análisis de enriquecimiento KEGG para el conjunto de genes expresados diferencialmente. Empleamos la función **enrichKEGG** que nos devolverá las categorías KEGG de enriquecimiento con control FDR.

```
kk_act <- enrichKEGG((act_NA_sorted), organism = "mouse", pvalueCutoff = 0.05)
```

El análisis de **enrichKEGG** señala que los DEGs activados están en los KEGG de enfermedades: alcoholismo y lupus eritematoso sistémico (Fig. 47).

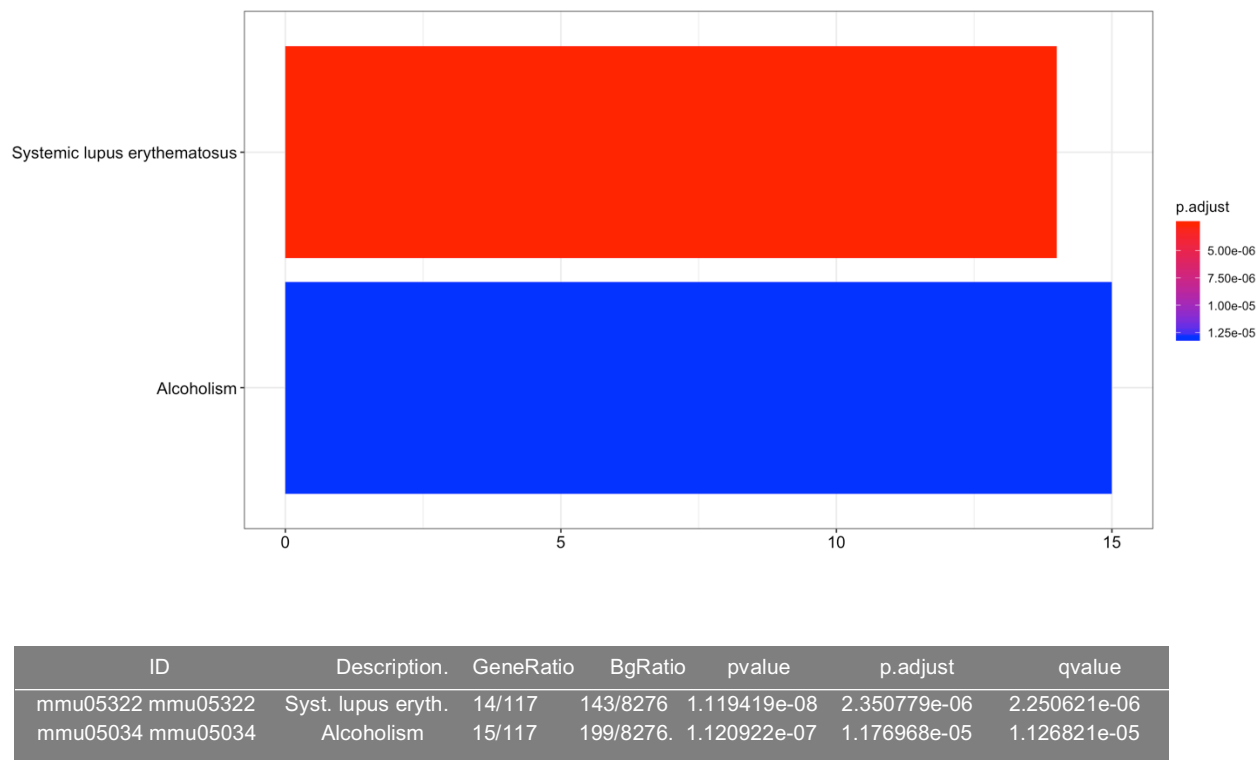
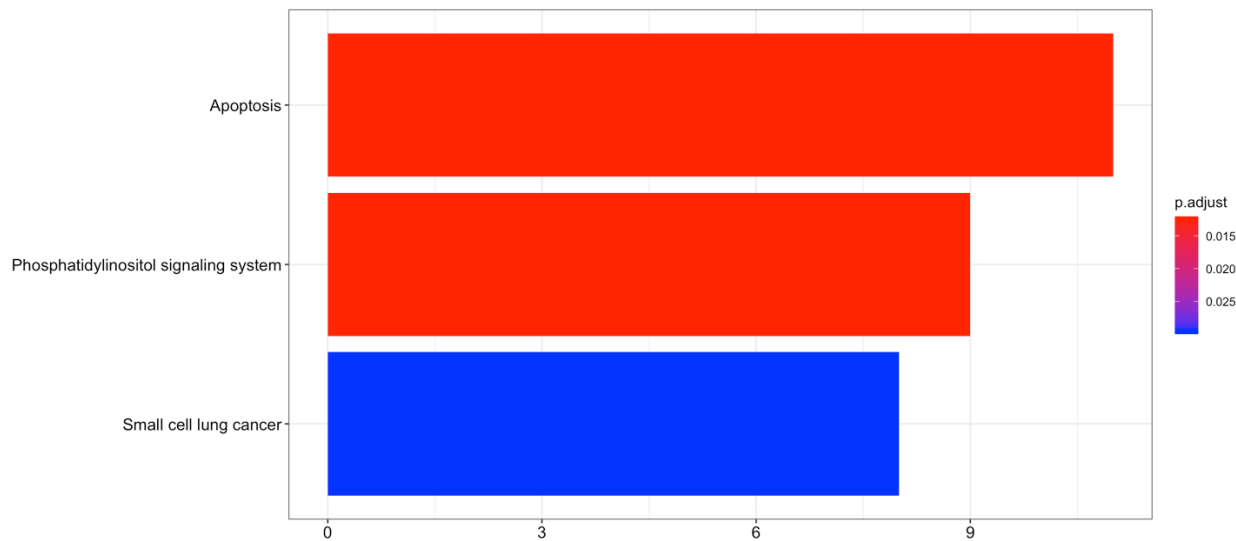


Fig. 47- barplot de enrichKEGG DEGs activados

Al realizar el análisis de **enrichKEGG** para los DEGs inhibidos obtenemos los KEGG: apoptosis, sistema de señalización fosfatidilinositol y cáncer pulmonar de células pequeñas (Fig. 48).



ID	Description.	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
mmu04210	mmu04210 Apoptosis	11/161	136/8276	0.0000642894	0.01360654	0.01185533
mmu04070	mmu04070 Phosp.inos sig.sys.	9/161	98/8276	0.0001143407	0.01360654	0.01185533
mmu05222	mmu05222 Small cell lung cancer	8/161	92/8276	0.0004014333	0.03184705	0.02774820

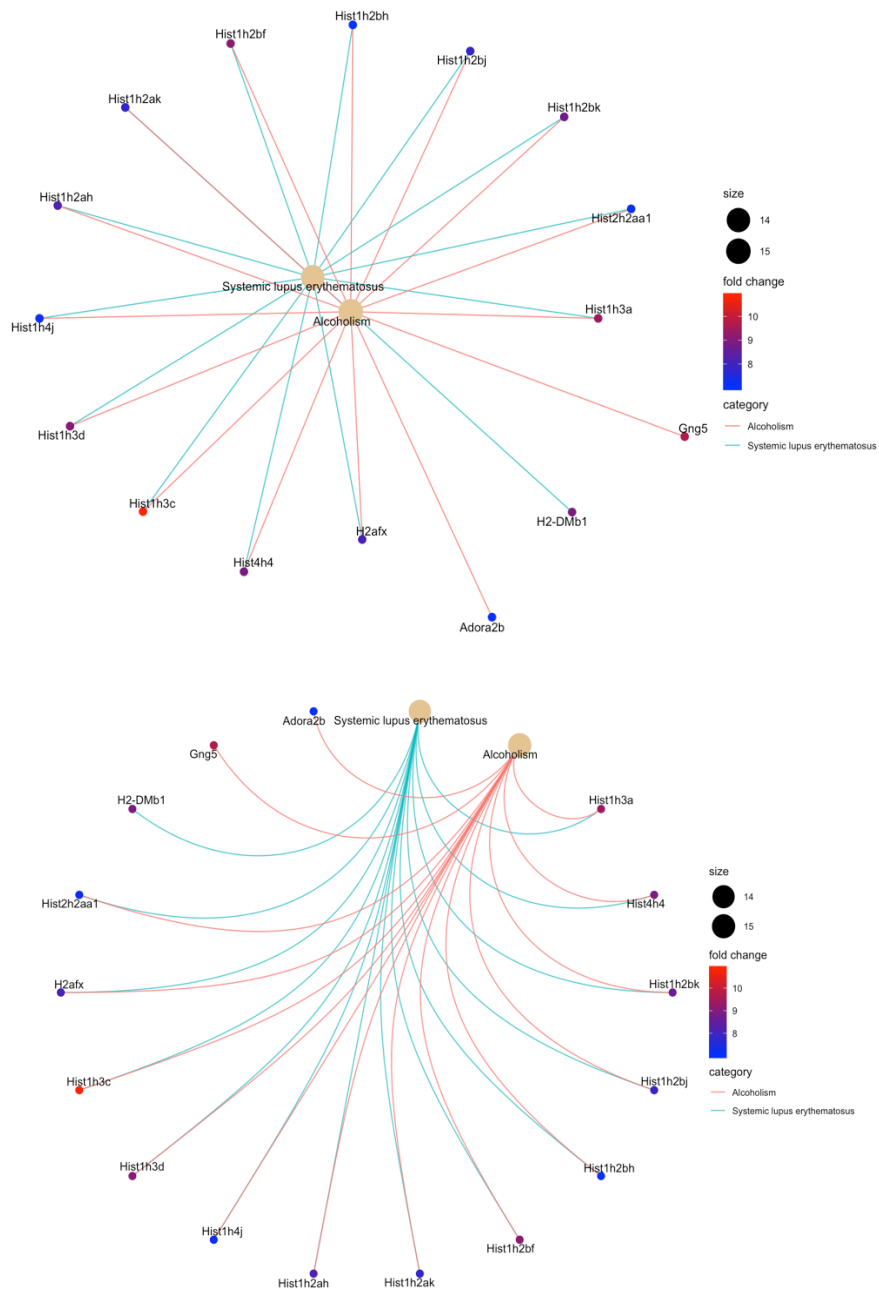
Fig. 48- barplot de enrcihKEGG DEGs inhibidos

Con el fin de poder saber y observar de forma más clara qué genes son los responsables de los enriquecimientos observados con anterioridad en KEGG; realizamos los **cnetplot** de los **enrichKEGG** obtenidos:

```
cnetplot(kk_act, foldChange = geneList_act2, colorEdge = TRUE)

cnetplot(kk_act, foldChange = geneList_act2, circular = TRUE, colorEdge = TRUE)
```

Mediante la representación de la forma clásica de **cnetplot** para KEGG se observa que el alcoholismo y lupus eritematoso sistémico comparten la mayor parte de los DEGs, entre los que se encuentran Hist1h3c, Gng5, Hist1h3d, Hist1h2bf, Hist1h3a, (Fig. 49, superior). La versión circular de **cnetplot** muestra que los únicos DEGs no compartidos son: alcoholismo (Adora2b, Gng5) y lupus eritematoso sistémico (H2-DMb1) (Fig. 49, inferior).



summary(kk_act)

ID	Description.	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
mmu05322	Syst. lupus erythem.	14/119	143/8222	1.516256e-08.	3.168976e-06	3.032513e-06
mmu05034	Alcoholism	15/119	199/8222	1.531449e-07	1.600364e-05	1.531449e-05

geneID

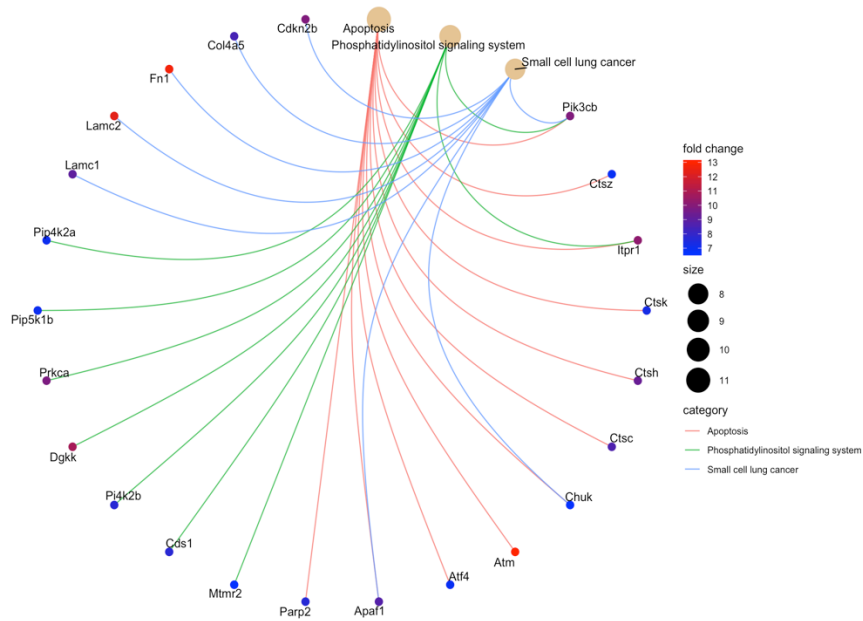
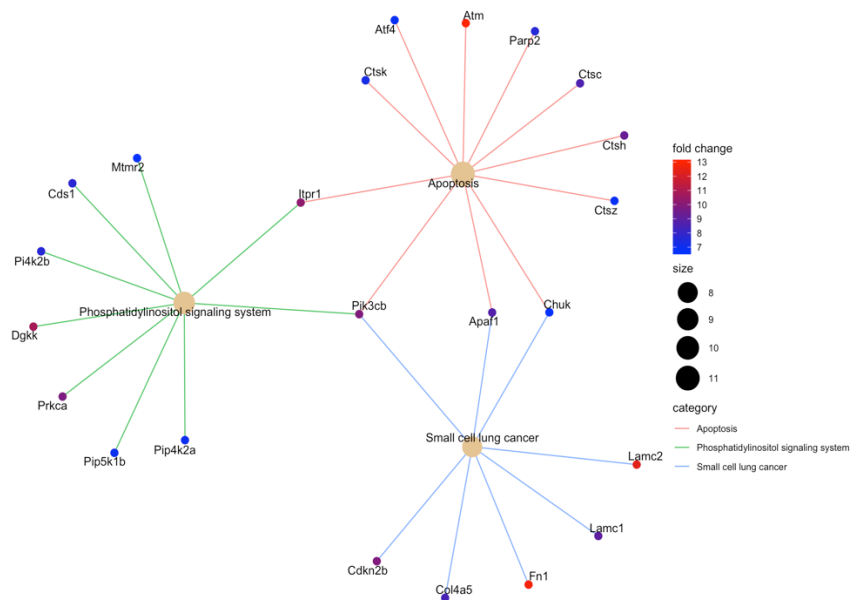
mmu05322

Hist1h3a/Hist4h4/Hist1h2bk/Hist1h2bj/Hist1h2bh/Hist1h2bf/Hist1h2ak/Hist1h2ah/Hist1h4j/Hist1h3d/Hist1h3c/H2afx/Hist2h2aa1/H2-DMb1

mmu05034

Hist1h3a/Hist4h4/Hist1h2bk/Hist1h2bj/Hist1h2bh/Hist1h2bf/Hist1h2ak/Hist1h2ah/Hist1h4j/Hist1h3d/Hist1h3c/H2afx/Hist2h2aa1/Gng5/Adora2b

Fig. 49- cnetplot de enciKEGG DEGs activados



summary(kk_inh)							
ID	Description.	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	
mmu04210	mmu04210 Apoptosis	11/161	136/8222	6.814183e-05	0.01417225	0.01245287	
mmu04070	mmu04070 Phos.inos. sign. Sys.	9/161	98/8222	1.201038e-04	0.01417225	0.01245287	
mmu05222	mmu05222 Small cell lung cancer	8/161	92/8222	4.191800e-04	0.03297549	0.02897490	
geneID							
mmu04210 Pik3cb/Ctsz/Itp1/Ctsk/Ctsh/Ctsc/Chuk/Atm/Atf4/Apaf1/Parp2							
mmu04070 Mtmr2/Pik3cb/Cds1/Pi4k2b/Dgkk/Prkca/Pip5k1b/Pip4k2a/Itp1							
mmu05222 Pik3cb/Lamc1/Lamc2/Fn1/Col4a5/Chuk/Cdkn2b/Apaf1							

Fig. 50- cnetplot de enrchiKEGG DEGs inhibidos

Mediante la representación de la forma clásica de **cnetplot** para KEGG observamos que los tres términos enriquecidos comparten pocos DEGs. Para apoptosis encontramos Atm, Itpr1, Pik3cb entre otros; sistema de señalización fosfatidilinositol, Dgkk, Prkca, Itpr1; cáncer pulmonar de células pequeñas, Lamc2, Fn1, Cdkn2b, (Fig. 50, superior). La versión circular de **cnetplot** muestra que los únicos DEGs compartidos son: apoptosis, cáncer pulmonar de células pequeñas, sistema de señalización fosfatidilinositol (Pik3cb); apoptosis, sistema de señalización fosfatidilinositol (Itpr1); apoptosis, cáncer pulmonar de células pequeñas (Apaf1, Chuk), (Fig. 50, inferior).

3.2.2.1- gseKEGG

Realizamos el análisis de enriquecimiento de conjuntos de genes de KEGG de los DEGs activados e inhibidos con la función **gseKEGG** con un `pvalueCutoff = 0.05`.

```
kk_act2 <- gseKEGG(geneList = geneList_act2, organism = "mouse", pvalueCutoff = 0.05)
```

No obtuvimos en ningún caso términos gseKEGG, tanto para los DEGs activados como inhibidos.

3.2.3- Reactome pathway database análisis de los genes expresados diferencialmente

Utilizamos el paquete de bioconductor ReactomePA que proporciona funciones para el análisis de rutas basadas en la base de datos de rutas REACTOME, implementando análisis de enriquecimiento y análisis de enriquecimiento de conjuntos de genes. La unidad central del modelo de datos Reactome es la reacción, las entidades que participan en las reacciones (ácidos nucleicos, proteínas, complejos, vacunas, terapias anticancerosas y moléculas pequeñas) forman una red de interacciones biológicas y se agrupan en vías. Los ejemplos de vías biológicas en Reactome incluyen el metabolismo intermediario clásico, la señalización, la regulación transcripcional, la apoptosis y la enfermedad. Realizamos el análisis de los genes expresados diferencialmente, activados e inhibidos de forma independiente.

3.2.3.1- ReactomePA

Realizamos el análisis del enriquecimiento de la vía de un conjunto de genes, dado un vector de genes, la función devolverá las rutas enriquecidas con control FDR.

Empleamos la función **enrichPathway** que nos devolverá las rutas enriquecidas con control FDR.

```
enr_P_act <- enrichPathway (geneList_actK, organism = "mouse", pvalueCutoff=0.05, pAdjustMethod = "BH", readable = FALSE)
```

El análisis de **enrichPathway** señala que los DEGs activados están en las rutas enriquecidas que se detallan a continuación: recombinación meiótica, regulación negativa de expresión rRNA por norC, sinapsis meiótica, metilación argininas de histonas RMTs, expresión de rRNA (por regulación epigenética positiva, regulación positiva del complejo B-WICH, regulación negativa SIRT1), entre otras, (Fig. 51). No obtuvimos ningún resultado de **enrichPathway** para los DEGs inhibidos.

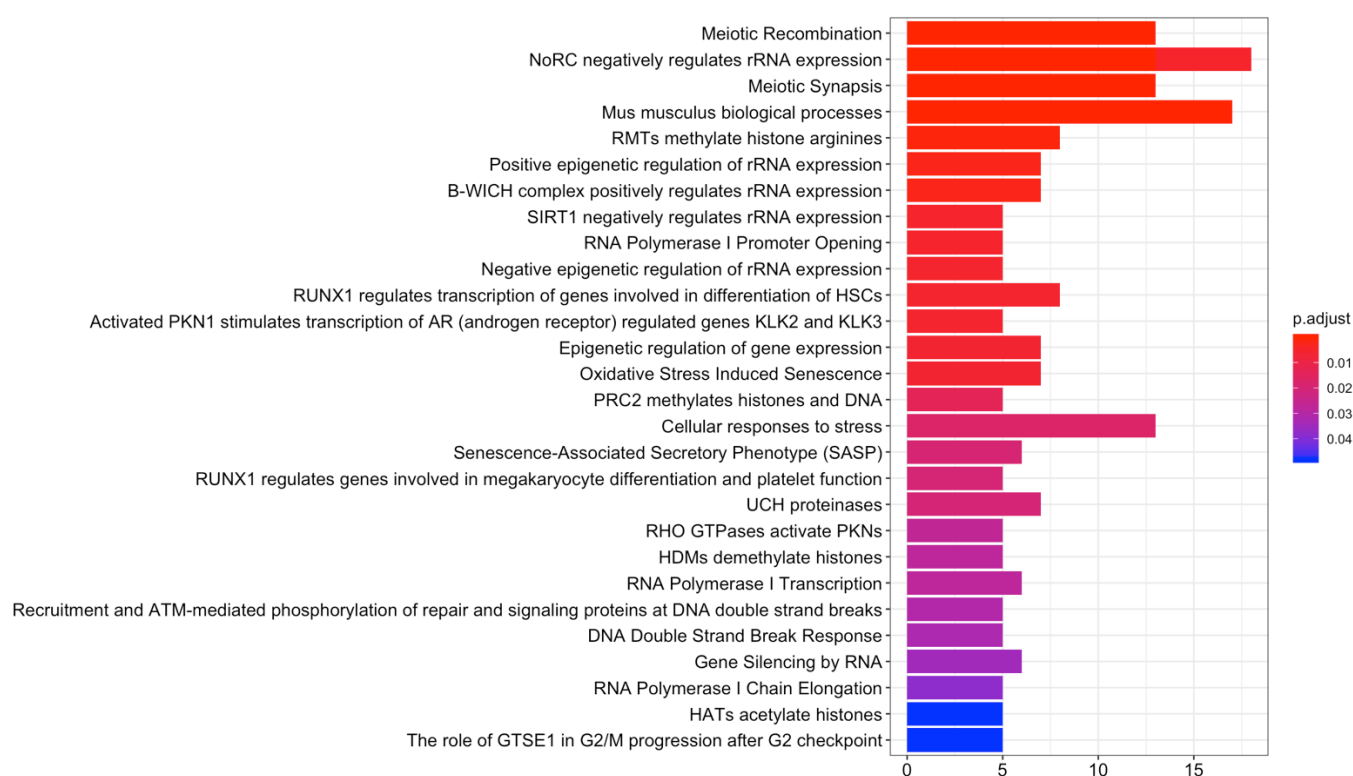


Fig. 51- barplot de encrichPathway DEGs activados

Mediante la representación de **cnetplot** para **enrichPathway** se observa que los DEGs mayoritarios son histonas, Hist1h2ak, Hist1h2ah, Hist1h3d, H2afx entre otras; las rutas obtenidas son; regulación negativa de expresión rRNA por norC, recombinación meiótica, sinapsis meiótica, metilación argininas de histonas RMTs. Encontramos además otros DEGs que tienen importancia en las rutas, Prmt7 (metilación argininas de histonas RMTs); Rps27a, Med11, Stat3, Bhlhe40 (procesos biológicos mus musculus), (Fig. 52).

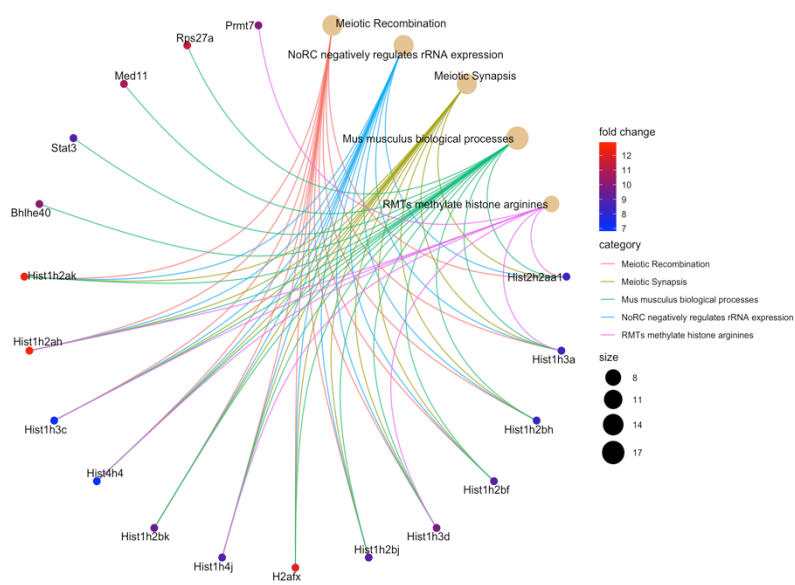
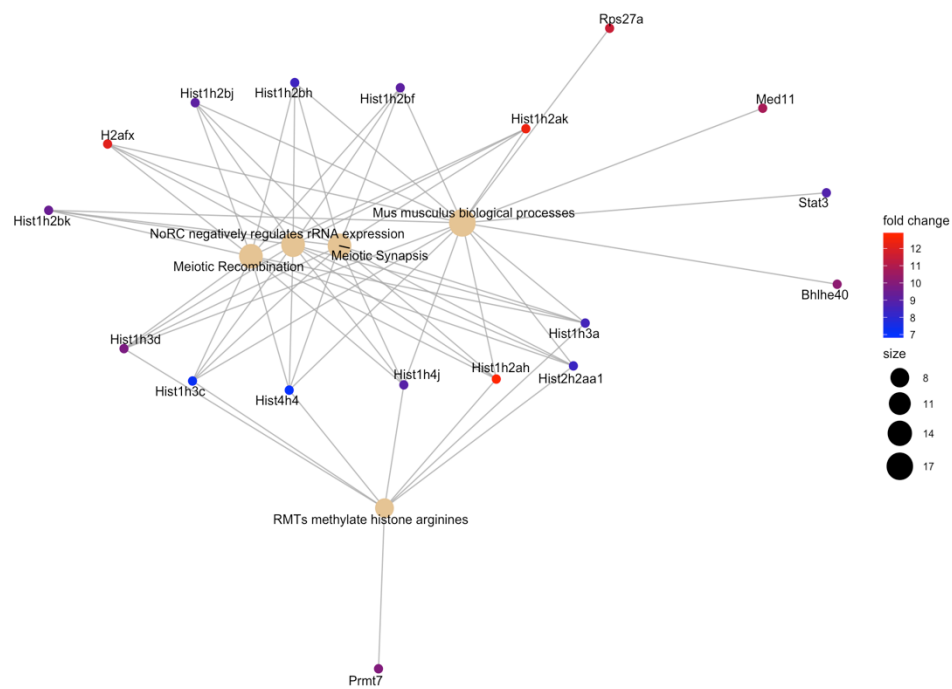


Fig. 52- cnetplot de encihPathway DEGs activados

4-DISCUSIÓN

Durante los últimos años diversos estudios han puesto de manifiesto la importante función que desempeña la proteína Ambra1 en el control de la autofagia y específicamente en la regulación de las primeras etapas de este proceso. Por otro lado, la desregulación de Ambra1 también parece jugar un papel en cáncer. En ese sentido, datos previos del grupo donde se ha desarrollado el presente trabajo indican que la pérdida de Ambra1 se asocia con una progresión maligna acelerada de carcinomas escamosos de piel en modelos animales asociada además con una aceleración de la transición epitelio mesénquima.

La búsqueda en bases de datos realizada en el TFM anterior ([Mendiburu-Eliçabe M., 2018](#)) apuntaba a la existencia de una posible relación entre una baja expresión de Ambra1 y peor pronóstico en pacientes de cáncer renal, pulmonar y gástrico. En cualquier caso, al tratarse de análisis realizados con bases de datos que contenían un número reducido de entradas derivadas de muestras metastáticas no fue posible inferir a partir de esos análisis que la expresión de Ambra1 estuviese asociada con el desarrollo de metástasis. Nuestra hipótesis de partida era que las muestras con baja expresión de Ambra1 tienen una mayor probabilidad de metastatizar, y aunque observamos esa posible tendencia en el análisis realizado, los datos no permitieron demostrarlo de una manera clara. Por otra parte, la funcionalidad de Ambra1 viene determinada en parte por las modificaciones post-traduccionales que sufre la proteína, tales como ubiquitinación y fosforilación ([Abrahamsen H., 2012](#); [Botti-Millet J., 2016](#)). Así, los cambios en los niveles/expresión de Ambra1 no serían el único evento que determina la existencia de alteraciones en la actividad de la proteína.

En este estudio, realizamos un código en R para el análisis de un array obtenido con Illumina. El array consta de un total de 6 muestras que corresponde a 2 grupos diferentes (3 muestras por grupo). El primer grupo “wild_type” (wt) son fibroblastos transformados. El segundo grupo “Ambra1_KO” (KO) son fibroblastos transformados deficientes en Ambra1.

En primer lugar, realizamos pre-procesado y control de calidad de las muestras, normalizamos por quantiles, transformamos a log2 con una varianza de corte igual a 0.5. Posteriormente realizamos el análisis de la expresión diferencial obteniendo aquellos genes que son significativos, los activados y los inhibidos.

Cuando comparamos el grupo wt con el KO, encontramos 4574 genes diferencialmente expresados con cambios significativos, 375 activados y 504 inhibidos. Como era de esperar Ambra1 es uno de los genes que se encuentran inhibidos en las

muestras KO, (Fig. 14), (Mendiburu-Eliçabe M., 2018). Sin embargo, no encontramos cambios significativos en algunos genes que habíamos seleccionado como potencialmente regulados (o implicados en la regulación de) Ambra1 en base a los resultados previos del grupo (TGF β 1, TGF β 2, TGF β 3, MYC, SQSTM1, ATG5 y ATG7). Existen diversas razones que podrían explicar que no se encuentren cambios en la expresión de esos genes en estas condiciones, incluyendo el tipo celular en el que se realizó el análisis (fibroblastos embrionarios transformados con Ras/E1A), puesto que tenemos evidencias experimentales anteriores de que la variación de Ambra1 promueve la des-fosforilación y degradación de c-MYC (Cianfanelli, V. et al., 2015).

En cualquier caso, aunque el análisis del array no permitió apoyar la hipótesis de partida, nos planteamos analizar las vías en las que están implicados los genes que se encuentran expresados diferencialmente (DEGs) entre el wt y el KO utilizando GO (Gene Ontology, Ashburner et al., 2000) y otras plataformas similares que contienen anotaciones acerca de los procesos biológicos, funciones moleculares o componentes celulares que caracterizan a grupos de genes, KEGG (Kanehisa, Goto, 2000).

En un experimento típico, de análisis de expresión génica mediante arrays, se generan listas de miles de genes de una colección de muestras pertenecientes a una de dos clases, por ejemplo, células wild type frente a modificadas. Los genes se pueden ordenar en una lista clasificada, de acuerdo con su expresión diferencial entre las clases. El reto es extraer el significado de esta lista. Un enfoque común consiste en centrarse en un puñado de genes en la parte superior e inferior de la lista obtenida (es decir, aquellos que muestran la mayor diferencia) para discernir pistas biológicas reveladoras. Este enfoque tiene algunas limitaciones importantes; después de corregir la prueba de múltiples hipótesis, ningún gen individual puede alcanzar el umbral de significación estadística, porque las diferencias biológicas relevantes son modestas en relación con el ruido inherente a la tecnología de microarrays. Alternativamente, uno puede quedarse con una larga lista de genes estadísticamente significativos sin ningún patrón biológico unificador, aunque en ese caso la interpretación puede ser tediosa. El análisis de un solo gen puede pasar por alto efectos importantes en las vías estudiadas ya que los procesos celulares a menudo afectan conjuntos de genes que actúan de manera coordinada. Así, un aumento del 20% en todos los genes que codifican enzimas de una ruta metabólica puede alterar dramáticamente el flujo a través de dicha ruta y puede ser más importante que un aumento de 20 veces en los niveles de un único gen. Por otra parte, cuando diferentes grupos estudian el mismo sistema biológico, la lista de genes con una expresión diferencial estadísticamente significativa entre los estudios desarrollados por esos grupos frecuentemente muestra un grado de concordancia sorprendentemente alto, lo que apoya la importancia de analizar grupos de genes

implicados en un determinado proceso más que genes individuales.

Por todas las razones descritas en el presente estudio nos planteamos ir un paso más allá y realizar el análisis de GO, KEGG y Reactome. De acuerdo con los resultados de la función de anotación de datos **basicProfile** a nivel 2, de la librería goProfiles, los DEGs se localizaron principalmente (CC) en orgánulos, membrana, complejo macromolecular y región extracelular. Se involucraron especialmente en procesos biológicos (BP) tales como organización de componentes celulares o biogénesis, regulación de procesos biológicos, procesos de desarrollo y respuesta a estímulo. Formaron parte de funciones moleculares (MF) tales como, “binding” y actividad catalítica, actividad reguladora transcripcional y regulación de función molecular. Encontramos diferencia significativa únicamente en los términos de BP entre los DEGs activados e inhibidos, observando que los procesos biológicos en los que hay diferencias son, proceso de desarrollo, localización, procesos de organismos multicelulares, proceso biológico de regulación positiva y señalización entre otras. Obtuvimos resultados muy similares a los conseguidos con el paquete goProfiles al utilizar el paquete clusterProfiler. En este caso utilizamos la función **groupGO** a nivel 3, y obtuvimos los términos GO para CC, orgánulos, membrana y región extracelular. Los procesos biológicos observados fueron, procesos metabólicos, regulación de procesos biológicos, procesos metabólicos de compuestos de nitrógeno, regulación procesos celulares, organización procesos celulares y procesos biosintéticos entre otros. Destacan por otra parte las funciones moleculares (MF) de binding tanto de proteínas como de compuestos orgánicos cíclicos, de compuestos heterocíclicos así como de iones.

Con las funciones **basicProfile** de la librería goProfiles y **groupGO** del paquete clusterProfiler pudimos obtener una primera estimación acerca de: (i) cuáles son los componentes celulares donde los DEGs están localizados (principalmente, orgánulos, membrana y región extracelular), (ii) cuales son los procesos biológicos en los que esos DEGs están implicados (procesos metabólicos y regulación de procesos biológicos entre otros); y (iii) cuales son las funciones moleculares que podrían estar influidas por los DEGs (principalmente de “binding” tanto de proteínas como compuestos orgánicos cíclicos y otros). Se han encontrado pocos estudios de arrays que analicen el papel de Ambra1, pero cabe destacar el análisis realizado en 25 conjuntos de datos de RNA-Seq de microarrays en cáncer de mama ([He RQ, et al., 2018](#)). En dicho estudio se obtienen resultados similares a los obtenidos en este trabajo ya que los autores encuentran cambios en componentes celulares enriquecidos (membrana y complejo macromolecular), los procesos biológicos de respuesta a estímulo, localización, regulación biológica y organización de componentes celulares y las funciones

moleculares de “binding”.

Para poder analizar cuáles de estos CC, BP y MF tienen mayor relevancia cuando inhibimos Ambra1 en fibroblastos, realizamos el cálculo de la función **enrichGO** del paquete clusterProfiler que incluye control mediante FDR, con un pValueCutoff = 0.05. En cuanto a los componentes celulares (CC), que se encontraron más significativos de los DEGs activados se localizaron: complejo de empaquetamiento de DNA (complejo de proteínas que desempeña un papel en el proceso de empaquetamiento del DNA) en la que se encuentran varias histonas (Hist1h2ak, Hist1h2bh, Hist1h1c); nucleosoma (complejo compuesto por DNA enrollado alrededor de un núcleo de varias subunidades y proteínas asociadas, que forma la unidad primaria de empaquetamiento del DNA en estructuras de orden superior) donde se encuentran también histonas (Hist1h2ah, Hist1h1c, H2afx); complejo proteína-DNA (complejo macromolecular que contiene proteínas y moléculas de DNA) donde se localizan histonas (Hist1h1c, Hist1h3a, Hist1h2bh); cromatina nuclear (complejo ordenado y organizado de DNA, proteínas y, a veces, RNA, que forma el cromosoma en el núcleo) donde se localizan Trp53, Stat3, Actb; y corteza celular (región de la célula que se encuentra justo debajo de la membrana plasmática y con frecuencia, pero no siempre, contiene una red de filamentos de actina y proteínas asociadas) donde se encuentran Crip2, Eno1b, Pfn1.

Los procesos biológicos (BP) más significativos en los que se encuentran involucrados los DEGs activados son: ensamblaje de cromatina (ensamblaje del DNA, histonas y otras proteínas asociadas y, en ocasiones, el RNA, en la estructura de la cromatina; comenzando con la formación de la unidad básica, el nucleosoma, seguido de la organización de los nucleosomas en estructuras de orden superior, dando lugar a una organización compleja de dominios específicos dentro del núcleo); empaquetamiento de DNA (cualquier proceso en el cual el DNA y las proteínas asociadas se forman en una estructura compacta y ordenada); ensamblaje y desensamblaje de cromatina (formación o destrucción de estructuras de cromatina); ensamblaje de nucleosomas (agregación, disposición y unión del nucleosoma); cambio conformación DNA (proceso celular que resulta en un cambio en la configuración espacial de una molécula de DNA. El cambio de conformación puede curvar el DNA o alterar el número de torsión, retorcimiento o el número de enlaces de una molécula de DNA). En todos estos procesos biológicos se encuentran involucradas varias histonas Hist1h1c, Hist1h2bf, además de Trp53, Cenpv, Ncaph2 entre otros.

Los resultados obtenidos para **enrichGO** tanto para componentes celulares como para procesos biológicos indican una fuerte relación entre la inhibición de Ambra1 y la activación de histonas implicadas en ensamblaje de cromatina, DNA-proteína,

empaquetamiento de DNA y organización del nucleosoma. Es interesante resaltar que esta relación entre Ambra1 y estabilidad del DNA no había sido documentada con anterioridad. Si existen trabajos en los que se relaciona Beclin-1 con la estabilidad del DNA, la mejora en la reparación del DNA y el refuerzo de la integridad general del genoma del DNA a través de la interacción con la Topoisomerasa-II β (Xu et al., 2017), (recordemos que Ambra1 es un componente clave del complejo Beclin1; Fig. 4). Para que se produzca esta interacción, Beclin-1 debe translocarse al núcleo, un evento que parece ocurrir de manera relativamente temprana durante el desarrollo murino (entre los días 15 y 20 postnatal; Xu et al., 2017).

Las funciones moleculares (MF) que encontramos enriquecidas de forma significativa en las células deficientes en Ambra1 fueron: actividad liasa carbono-azufre (eliminan ácido sulfhídrico, H₂S, o derivados sustituidos), DEGs implicados Cenpv, Glo1, Mgst3, Kyat3; unión de cadherinas (interactuando de manera selectiva y no covalente con la cadherina, una proteína de membrana de tipo I involucrada en la adhesión celular); unión de la molécula de adhesión celular (interactuando selectivamente y no covalentemente con una molécula de adhesión celular). Los DEGs implicados en estas dos funciones moleculares son las histonas Hist1h3d y Hist1h3a además de Eno1b y Pfn1 entre otros. En líneas con esas observaciones, recientemente se ha sido descrito el papel de Ambra1 en adhesión en células de carcinoma escamoso de ratón mediante el control de la ruta Src/FAK. Así, cuando Ambra1 no puede unirse a FAK, los niveles anormalmente altos de fosfo-Src y fosfo-FAK se acumulan en las adherencias focales, regulando positivamente la adhesión y la migración invasiva (Schoenherr C. et al, 2017).

Cuando aplicamos la función **enrichGO** sobre los DEGs inhibidos encontramos que éstos estaban enriquecidos, de forma significativa, en los componentes celulares (CC): (i) parte apical de la célula (región de la célula polarizada que forma una punta o es distal a una base. Por ejemplo, en una célula epitelial polarizada, la región apical tiene una superficie expuesta y se encuentra opuesta a la lámina basal que separa el epitelio de otro tejido); (ii) membrana plasmática apical (región de la membrana plasmática ubicada en el extremo apical de la célula) (hay que destacar que en estos dos CC se encuentran los DEGs Fat1, Slc12a2, Shroom2 y Hsp90ab1); (iii) membrana vacuola (bicapa lipídica que rodea la vacuola y separa su contenido del citoplasma de la célula); (iv) vacuola lítica (vacuola que se mantiene a un pH ácido y que contiene enzimas degradativas, incluida una amplia variedad de hidrolasas ácidas); y (v) lisosoma (pequeña vacuola lítica que tiene una morfología independiente del ciclo celular y se encuentra en la mayoría de las células animales y que contiene una variedad de hidrolasas, la mayoría de las cuales tienen sus actividades máximas en el rango de

pH 5-6) (los DEGs de estos CC son Hsp90ab1, Vps35, Ahnak, entre otros).

Los procesos biológicos (BP) que encontramos enriquecidos de forma significativa para los DEGs inhibidos fueron: (i) estructura anatómica de homeostasis (proceso homeostático involucrado en el mantenimiento de estado estacionario interno dentro de una estructura anatómica definida de un organismo, incluido el control de la proliferación celular y la muerte y el control de la función metabólica) (en ese sentido cabe destacar los cambios encontrados en los niveles de Sox9, Gcnt2 y Lapm2); (ii) respuesta a transformación de crecimiento factor beta (cualquier proceso que resulte en un cambio en el estado o actividad de una célula o un organismo, en términos de movimiento, secreción, producción de enzimas, expresión génica, como resultado de un estímulo beta del factor de crecimiento transformante) (de nuevo con la existencia de cambios en Sox9, Hsp90ab1 y Htra1); (iii) morfogénesis del tubo (proceso en el que se generan y organizan las estructuras anatómicas de un tubo; los tubos epiteliales y endoteliales transportan gases, líquidos y células de un sitio a otro y forman la estructura básica de muchos órganos y tejidos, con la forma y la organización de los tubos que varían de una a otra) (con la existencia de cambios en los genes Sox9, Slc12a2, Stk4); (iv) regulación positiva de la motilidad celular (cualquier proceso que active o aumente la frecuencia, velocidad o extensión de la motilidad celular) (con cambios en la expresión de Sox9, Gcnt2, Ets1); (v) respuesta a partículas lipoproteicas (cualquier proceso que resulte en un cambio en el estado o actividad de una célula o un organismo, en términos de movimiento, secreción, producción de enzimas, expresión génica, etc., como resultado de un estímulo de partículas de lipoproteínas) (con cambios en la expresión de Npc1, Lpl y Adam17). En relación con estas observaciones cabe destacar que la expresión de Ambra1 se asoció positivamente con la expresión de Snail, uno de los principales factores transcripcionales implicados en la regulación de la transición epitelio mesénquima (EMT), un proceso que se inicia con la pérdida de la polaridad apical-basal a medida que se pierden las uniones estrechas, permitiendo que se entremezclen los componentes de las membranas apical y baso-lateral, acompañada por disminución de la expresión de claudinas y ocludinas, adquisición de morfología de células fusiformes y reorganización del citoesqueleto, (Nitta T, et al, 2014). Así un estudio previo correlacionaba la sobreexpresión de Ambra1 con metástasis en ganglios linfáticos y baja tasa de supervivencia en pacientes con colangiocarcinoma, lo que llevaba a los autores a sugerir que la autofagia puede asociarse con un fenotipo maligno y un pronóstico desfavorable, mediante la regulación de la EMT (Takeo N. et al., 2014) .

La función molecular (MF) más significativa en la que se encuentran involucrados los DEGs inhibidos es la actividad transferasa de grupos glicosilo (catálisis de la transferencia de un grupo glicosilo de un compuesto donante a otro aceptor). Siendo los

DEGs implicados en ese proceso Gcnt2, Ugcg y Tnks2.

Realizamos el análisis de **enrichKEGG** de los DEGs obtenidos cuando inhibimos Ambra1 en fibroblastos, del paquete clusterProfiler que incluye control mediante FDR, con un pValueCutoff = 0.05. Las categorías de KEGG enriquecidas para los DEGs activados son lupus eritematoso sistémico (enfermedad autoinmune prototípica caracterizada por la producción de autoanticuerpos IgG específicos para antígenos propios, como el DNA, las proteínas nucleares y ciertos componentes citoplásmicos, en asociación con una amplia variedad de manifestaciones clínicas) y alcoholismo, los DEGs que se encontraron involucrados en estas dos enfermedades fueron, Hist1h3c, Hist1h3a, Hist1h3d entre otros.

Los KEGG enriquecidos en nuestro análisis para los DEGs inhibidos son, (i) apoptosis (proceso programado genéticamente para la eliminación de células dañadas o redundantes mediante la activación de caspasas) los DEGs asociados son Atm, Pik3cb, Itpr; (ii) sistema de señalización de fosfatidilinositol, los DEGs que se encontraron involucrados fueron, Dgkk, Pik3cb, Prkca; (iii) carcinoma de pulmón de células pequeñas (es una neoplasia muy agresiva, que representa aproximadamente el 25% de todos los casos de cáncer de pulmón. Los mecanismos moleculares alterados en el CPCP incluyen la expresión inducida de oncogén, MYC y la pérdida de genes supresores de tumores, como p53, PTEN, RB y FHIT), los DEGs asociados son Fn1, Lamc2, Pik3cb. La función de Ambra1 en la autofagia y la apoptosis se ha explorado in vitro en células madre embrionarias y fibroblastos humanos. Así, entre otros estudios el de [Gu W., et al., \(2014\)](#) sugiere que Ambra1 funciona en la intersección entre la autofagia y la apoptosis; Ambra1 interactúa con Beclin1 para funcionar como un interruptor pro-supervivencia que inhibe la apoptosis e induce la autofagia, previniendo así la muerte de las células cancerígenas colorrectales en respuesta al tratamiento con agentes pro-apoptóticos.

En último lugar, el análisis de **enrichPathway** del paquete ReactomePA, sólo nos dio rutas enriquecidas para los DEGs activados, (i) recombinación meiótica (intercambia segmentos de ADN dúplex entre homólogos cromosómicos, generando diversidad genética), los DEGs asociados son Hist1h3d, H2afx, Hist1h2bk; (ii) regulación negativa de expresión rRNA por norC (el complejo de remodelación nuclear (NoRC), compuesto por TIP5 (BAZ2A) y el remodelador de cromatina SNF2H (SMARCA5), silencia el gen rRNA), DEGs que se encontraron involucrados fueron H2afx, Hist1h2ah, Hist1h3d; (iii) sinapsis meiótica (es la pareja física estable de cromosomas homólogos que comienza en el leptonema de la profase I y dura hasta la anafase de la profase I), los DEGs asociados son Hist1h2ah, Hist1h2ak, H2afx; (iv) metilación argininas de histonas RMTs (modificación postraduccional común; alrededor

del 2% de los residuos de arginina se metilan en núcleos de hígado de rata), los DEGs involucrados son Prmt7, Hist1h2ah, Hist1h3d.

Los resultados obtenidos en el presente TFM confirman la hipótesis inicial de la existencia de una relación entre inhibición de Ambra1 y adhesión molecular, pérdida de la polaridad apical-basal en el proceso EMT, apoptosis, señalización fosfatidilinositol, lisosoma, confirmando de esta manera que la inhibición de Ambra1 en los fibroblastos murinos fue realizada con éxito. La novedad del estudio radica en la relación encontrada entre la inhibición de Ambra1 y la activación de histonas tales como, Hist1h2ah, H2afx, Hist1h1c, entre otras que se encuentran implicadas en diversos procesos biológicos como ensamblaje de cromatina, ensamblaje de nucleosoma, empaquetamiento de DNA, ensamblaje/desensamblaje de cromatina, recombinación meiótica, expresión de rRNA que no ha sido descrito con anterioridad. Estos resultados podrían explicar además unos resultados obtenidos con anterioridad en el laboratorio donde se desarrolló el TFM en el contexto de una colaboración con el grupo de Francesco Cecconi (Copenhague, Dinamarca), en los cuales encontramos que el fármaco AZD7762, potente inhibidor de la quinasa de punto de control competitivo de ATP es más efectivo en los MEFs (Mouse Embryonic Fibroblast) Ras E1A deficientes en AMBRA1 que en los wild type (WT) (Fig. 53). Además, la inhibición observada en los DEGs de la ruta de señalización de fosfatidilinositol también podría ser un testigo de la existencia de alteraciones en el ciclo celular y en la reparación de DNA, lo que también podría contribuir a explicar la mayor efectividad del AZS7762 en células deficientes en AMBRA1 (Zabludoff SD., et al, 2008).

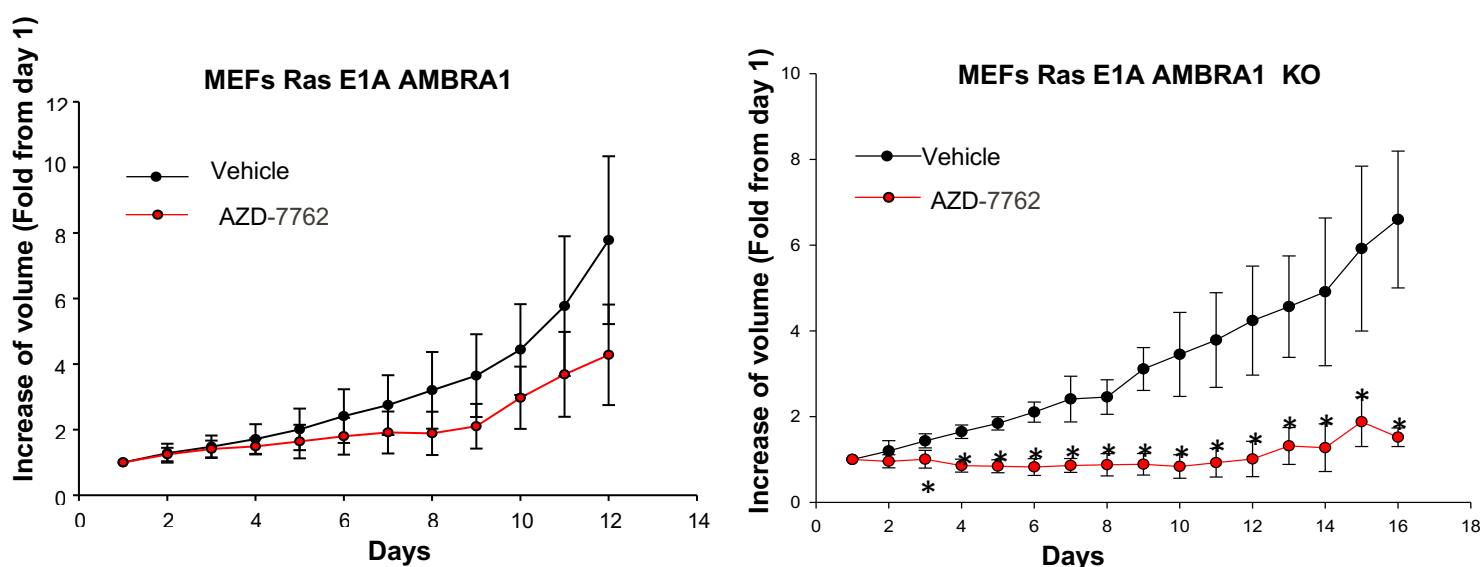


Fig. 53- MEFs Ras E1A AMBRA1 y MEFs Ras E1A AMBRA1 KO tratados con AZD-7762

Con el fin de confirmar la relevancia de estas observaciones en pacientes sería interesante desarrollar una búsqueda que permitiera analizar el perfil de DEGs en estudios de pacientes con cáncer depositados en GEO, ArrayExpress, Oncomine, o TCGA y en los que se haya analizado grado de expresión de AMBRA1.

5- CONCLUSIONES

- El análisis del perfil de expresión génica de fibroblastos embrionarios transformados con Ras/E1A con genotipo “wild type” o deficientes en Ambra1, indica que existen genes diferencialmente expresados (DEGs) entre el grupo wt y el KO; 375 activados y 504 inhibidos.
- En el análisis **basicProfile** de los DEGs se obtuvieron los CC: membrana, complejo macromolecular y región extracelular. Se involucraron especialmente en BP tales como organización de componentes celulares o biogénesis, regulación de procesos biológicos, procesos de desarrollo y respuesta a estímulo. Formaron parte de FM tales como, “binding” y actividad catalítica, actividad reguladora transcripcional y regulación de función molecular.
- El análisis de la función **enrichGO** de los DEGs activados respecto a los CC reveló la existencia de los siguientes componentes: complejo de empaquetamiento de DNA, nucleosoma, complejo proteína-DNA, cromatina nuclear, y corteza celular. Los BP obtenidos fueron: ensamblaje de cromatina, empaquetamiento de DNA, ensamblaje y desensamblaje de cromatina y cambio conformación DNA. Las MF que encontramos enriquecidas fueron: actividad liasa carbono-azufre, unión de cadherinas, unión de la molécula de adhesión celular.
- La función **enrichGO** sobre los DEGs inhibidos resultaron en los CC enriquecidos: parte apical de la célula, membrana plasmática apical, vacuola lítica. Los BP que encontramos fueron: estructura anatómica de homeostasis, respuesta a transformación de crecimiento factor beta, morfogénesis del tubo, regulación positiva de la motilidad celular. La MF más significativa es actividad transferasa,
- El análisis de **enrichKEGG** para los DEGs inhibidos dio como resultado: apoptosis y sistema de señalización de fosfatidilinositol.
- Los resultados obtenidos en el presente TFM confirman la hipótesis inicial de inhibición de Ambra1 y adhesión molecular, pérdida de la polaridad apical-basal en el proceso EMT, apoptosis, señalización fosfatidilinositol y aporta como novedad la relación encontrada entre la inhibición de Ambra1 y la activación de histonas que implica diversos procesos biológicos como ensamblaje de cromatina, ensamblaje de nucleosoma, empaquetamiento de DNA, ensamblaje/desensamblaje de cromatina, recombinación meiótica, expresión de rRNA que no ha sido descrito con anterioridad.

6- BIBLIOGRAFÍA

Abrahamsen H, Stenmark H, Platta HW. Ubiquitination and phosphorylation of Beclin 1 and its binding partners: Tuning class III phosphatidylinositol 3-kinase activity and tumor suppression. *FEBS Lett.* 2012 Jun 4;586(11):1584-91. doi: 10.1016/j.febslet.2012.04.046. Epub 2012 May 3.

Alexa A, Rahnenfuhrer J. "topGO: Enrichment Analysis for Gene Ontology. R package version 2.34.0." (2018)

Ashburner et al. "Gene ontology: tool for the unification of biology" (2000) *NatGenet* **25**(1):25-9.

Ashburner, M. et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25(1), pp.25–29.

Benato et al., 2013, Ambra1 knockdown in zebrafish leads to incomplete development due to severe defects in organogenesis. *Autophagy*. 2013 Apr;9(4):476-95. doi: 10.4161/auto.23278. Epub 2013 Jan 24.

Bioconductor. <https://www.bioconductor.org>

Botti-Millet J., et al., Fine-tuning autophagy: from transcriptional to posttranslational regulation. *Am J Physiol Cell Physiol*. 2016 Sep 1;311(3):C351-62. doi: 10.1152/ajpcell.00129.2016. Epub 2016 Jun 22.

Carlson M, "KEGG.db: A set of annotation maps for KEGG. R package version 3.2.3."(2016).

Carlson M. GO.db: "A set of annotation maps describing the entire Gene Ontology. R package version 3.7.0." (2018)

Cecconi et al., 2008, The involvement of cell death and survival in neural tube defects: a distinct role for apoptosis and autophagy? *Cell Death Differ.* 2008 Jul;15(7):1170-7. doi: 10.1038/cdd.2008.64. Epub 2008 May 2.

Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. May 2012 2; 401.

Choi, A.M., Ryter, S.W. & Levine, B. Autophagy in human health and disease. *N Engl J Med* **368**, 1845-6 (2013).

Di Bartolomeo et al., 2010, The dynamic interaction of AMBRA1 with the dynein motor complex regulates mammalian autophagy. *J Cell Biol.* 2010 Oct 4;191(1):155-68. doi: 10.1083/jcb.201002100.

Dunning M, Lynch A and Eldridge M (2015). illuminaMousev2.db: Illumina MouseWG6v2 annotation data (chip illuminaMousev2). R package version 1.26.0.

Fabregat A., et al, "The Reactome Pathway Knowledgebase." *Nucleic Acids Res.* 2018 Jan 4;46(D1):D649-D655. doi: 10.1093/nar/gkx1132.

Fimia GM, Corazzari M, Antonioli M, Piacentini M (2013) Ambra1 at the crossroad between autophagy and cell death. *Oncogene* 32: 3311–3318.

Fimia GM, et al. Ambra1 regulates autophagy and development of the nervous system. *Nature*. 2007; 447:1121–5. [PubMed: 17589504]

Fimia y Piacentini, 2010, Regulation of autophagy in mammals and its interplay with apoptosis. *Cell Mol Life Sci.* 2010 May;67(10):1581-8. doi: 10.1007/s00018-010-0284-z. Epub 2010 Feb 18.

Forbes SA, et al., 2016, COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research*, Volume 45, Issue D1, 4 January 2017, Pages D777–D783, <https://doi.org/10.1093/nar/gkw1121>

Galluzzi L, et al., Autophagy in malignant transformation and cancer progression. *EMBO J.* 2015 Apr 1;34(7):856-80. doi: 10.15252/embj.201490784. Epub 2015 Feb 23. Review.

Gao et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, p1 (2013).

Gentleman et al., Bioconductor: open software development for computational biology and bioinformatics. 2004 *Genome Biology* 5:R80

ggplot2. <https://CRAN.R-project.org/package=ggplot2>

GO Consortium, *Nucleic Acids Res.*, 2017

ggplots. <https://CRAN.R-project.org/package=ggplots>

Gu W., et al., "Ambra1 is an essential regulator of autophagy and apoptosis in SW620 cells: pro-survival role of Ambra1". PLoS One. 2014 Feb 26;9(2):e90151. doi: 10.1371/journal.pone.0090151. eCollection 2014.

Hay ED. An overview of epithelio-mesenchymal transformation. Acta Anat. 1995; 154:8–20. [PubMed: 8714286]

He RQ1, et al. "The Clinicopathological Significance and Correlative Signaling Pathways of an Autophagy-Related Gene, Ambra1, in Breast Cancer: a Study of 25 Microarray RNA-Seq Datasets and in-House Gene Silencing" Cell Physiol Biochem. 2018;51(3):1027-1040. doi: 10.1159/000495483. Epub 2018 Nov 26.

Huber et al., Orchestrating high-throughput genomic analysis with Bioconductor. 2015 Nature Methods 12:115-121

illuminaMousev2.db, <http://bioconductor.org/packages/illuminaMousev2.db/>

Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009 Jun;119(6):1420-8. doi: 10.1172/JCI39104. Review. Erratum in: J Clin Invest. 2010 May 3;120(5):1786.

Kanehisa M., et al, "KEGG: new perspectives on genomes, pathways, diseases and drugs". Nucleic Acids Res. 2017 Jan 4;45(D1):D353-D361. doi: 10.1093/nar/gkw1092. Epub 2016 Nov 28.

Kanehisa M., et al, "New approach for understanding genome variations in KEGG". Nucleic Acids Res. 2018 Oct 13. doi: 10.1093/nar/gky962.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28(1), pp.27–30.

Kanehisa, M., Toward pathway engineering: a new database of genetic and molecular pathways. Science & Technology Japan, No. 59, pp. 34-38 (1996)

Kroemer G, Marino G, Levine B Autophagy and the integrated stress response. Mol Cell 40: 280 – 293, (2010)

Lanczky A, et al., miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2,178 breast cancer patients, Breast Cancer Res Treat. 2016;160(3):439-446.

Leland Wilkinson. The Grammar of Graphics. Statistics and Computing. Springer, 2nd edition, 2005.

Ligtenberg W., " reactome.db: A set of annotation maps for reactome. R package version 1.66.0." (2018).

LIMMA, Linear Models for Microarray Data <http://bioconductor.org/packages/release/bioc/html/limma.html>

Ma Y, et al. Autophagy and cellular immune responses. Immunity 39: 211 – 227, (2013)

Matloff N., The art of R programming. 2011 (Libro)

Mei, Y., Su, M., Soni, G., Salem, S., Colbert, C. L. and Sinha, S. C. (2014). Intrinsically disordered regions in autophagy proteins. Proteins 82, 565-578. doi:10.1002/prot.24424

Mendiburu-Eliçabe M., "Estudio bioinformático de la proteína reguladora de autofagia Ambra1 en cáncer", trabajo fin de máster Máster en bioinformática y biología computacional Escuela Nacional de Salud- Instituto de Salud Carlos III, (2018).

Michelle Cicchini, et al. Molecular Pathways: Autophagy in Cancer—A Matter of Timing and Context. DOI: 10.1158/1078-0432.CCR-13-2438 Published February 2015

Mizushima N, Komatsu M. Autophagy: renovation of cells and tissues. Cell. 2011 Nov 11;147(4):728-41. doi: 10.1016/j.cell.2011.10.026. Review.

Mizushima N., Autophagy: process and function. Genes Dev. 2007 Nov 15;21(22):2861-73. Review.

Mizushima, N., Levine, B., Cuervo, A.M. & Klionsky, D.J. Autophagy fights disease through cellular self-digestion. Nature 451, 1069-75 (2008).

Mootha V., et al, "PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes". Nature Genetics volume 34, pages 267–273 (2003).

Mootha, et al. "PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics volume 34, pages 267–273 (2003)

Morselli E, et al., Oncosuppressive functions of autophagy. Antioxid Redox Signal. 2011 Jun;14(11):2251-69. doi: 10.1089/ars.2010.3478. Epub 2010 Dec 4. Review.

National Cancer Institute, <https://visualsonline.cancer.gov/>

Nazio et al., 2013, mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. Nat Cell Biol. 2013 Apr;15(4):406-16. doi: 10.1038/ncb2708. Epub 2013 Mar 24.

- Nazio F**, Cecconi F., 2013, mTOR, AMBRA1, and autophagy: an intricate relationship. *Cell Cycle*. 2013 Aug 15;12(16):2524-5. doi: 10.4161/cc.25835. Epub 2013 Jul 30.
- Nazio F**, et al. mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. *Nat Cell Biol*. 2013; 15:406–16. [PubMed: 23524951]
- Nitta T**. et al. "Autophagy may promote carcinoma cell invasion and correlate with poor prognosis in holangiocarcinoma". *Int J Clin Exp Pathol*. 2014 Jul 15;7(8):4913-21. eCollection 2014.
- Parzych KR**. and Klionsky DJ. An Overview of Autophagy: Morphology, Mechanism, and Regulation. *Antioxid Redox Signal*. 2014 Jan 20;20(3):460-73. doi: 10.1089/ars.2013.5371. Epub 2013 Aug 2. Review.
- Ritchie ME**, et al., (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." *Nucleic Acids Research*, 43(7), pp. e47.
- RStudio**. <https://www.rstudio.com/products/RStudio/>
- Samy Lamouille**,¹ Jian Xu,² and Rik Derynck¹, *Nat Rev Mol Cell Biol*. 2014 Mar; 15(3): 178–196.
- Sanchez A**, et al., "goProfiles: goProfiles: an R package for the statistical analysis of functional profiles". R package version 1.44.0. (2018).
- Schoenherr C**. et al, "Ambra1 spatially regulates Src activity and Src/FAK-mediated cancer cell invasion via trafficking networks". *Elife*. 2017 Mar 31;6. pii: e23172. doi: 10.7554/eLife.23172.
- Stolz A**, Ernst A, Dikic I. Cargo recognition and trafficking in selective autophagy. *Nat Cell Biol*. 2014 Jun;16(6):495-501. doi: 10.1038/ncb2979. Review.
- Subramanian A**., et al, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". *Proc Natl Acad Sci U S A*. 2005 Oct 25;102(43):15545-50. Epub 2005 Sep 30.
- Subramanian**, et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles", *PNAS* October 25, 2005 102 (43) 15545-15550; published ahead of print September 30, 2005
- Takeo N**. et al, "Autophagy may promote carcinoma cell invasion and correlate with poor prognosis in cholangiocarcinoma". *Int J Clin Exp Pathol*. 2014; 7(8): 4913–4921. Published online 2014 Jul 15. PMID: PMC4152052. PMID: 25197362
- Tenenbaum D**., "KEGGREST: Client-side REST access to KEGG. R package version 1.22.0." (2018).
The Human Protein Atlas, <https://www.proteinatlas.org>
- Thiery JP**, et al., Epithelial-mesenchymal transitions in development and disease. *Cell*. 2009 Nov 25;139(5):871-90. doi: 10.1016/j.cell.2009.11.007. Review.
- Thiery JP**, Sleeman JP. Complex networks orchestrate epithelial-mesenchymal transitions. *Nature Rev.Mol.Cell Biol*. 2006 ;7:131–142.[PubMed:16493418]
- Thul PJ**, et al., A subcellular map of the human proteome. *Science*. 2017 356(6340): eaal3321. PubMed: 28495876. DOI: 10.1126/science.aal3321
- Uhlen M**, et al., A pathology atlas of the human cancer transcriptome. *Science*. 2017 357(6352). PubMed: 28818916. DOI: 10.1126/science.aan2507
- Uhlén M**, et al., Tissue-based map of the human proteome. *Science* 2015 347(6220):1260419. PubMed: 25613900 DOI: 10.1126/science.1260419
- Uhlen M**, et al., Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010 28(12):1248-50. PubMed: 21139605. DOI: 10.1038/nbt1210-1248
- Uniprot**, <http://www.uniprot.org>
- Valentina Cianfanelli**, et al. AMBRA1 links autophagy to cell proliferation and tumorigenesis by promoting c-MYC dephosphorylation and degradation. *Nat Cell Biol*. 2015 January ; 17(1): 20–30. doi:10.1038/ncb3072.
- Verfaillie T**., Salazar, M., Velasco, G. & Agostinis, P. Linking ER Stress to Autophagy: Potential Implications for Cancer Therapy. *Int J Cell Biol* **2010**, 930509 (2010).
- Warnes GR**. Et al., 2016, Various R Programming Tools for Plotting Data
- White, E**. Deconvoluting the context-dependent role for autophagy in cancer. *Nat Rev Cancer* **12**, 401-10 (2012).
- Wickham H**., and Chang W., 2016, Create Elegant Data Visualisations Using the Grammar of Graphics
- Xu F**, et al. "Nuclear localization of Beclin 1 promotes radiation-induced DNA damage repair independent of autophagy". *Sci Rep*. 2017 Mar 27;7:45385. doi: 10.1038/srep45385.

Yu G, et al., "clusterProfiler: an R package for comparing biological themes among gene clusters." *OMICS: A Journal of Integrative Biology*, 16(5), 284-287. doi: 10.1089/omi.2011.0118. (2012).

Yu G, et al., "DOSE: an R/Bioconductor package for Disease Ontology Semantic and Enrichment analysis." *Bioinformatics*, 31(4), 608-609. (2015) doi: 10.1093/bioinformatics/btu684, <http://bioinformatics.oxfordjournals.org/content/31/4/608>.

Yu G, He Q. "ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization." *Molecular BioSystems*, 12(12), 477-479. (2016). doi: 10.1039/C5MB00663E, <http://pubs.rsc.org/en/Content/ArticleLanding/2015/MB/C5MB00663E>

Yu G., "enrichplot: Visualization of Functional Enrichment Result. R package version 1.2.0", <https://github.com/GuangchuangYu/enrichplot>. (2018).

Zabludoff SD., "AZD7762, a novel checkpoint kinase inhibitor, drives checkpoint abrogation and potentiates DNA-targeted therapies". *Mol Cancer Ther*. 2008 Sep;7(9):2955-66. doi: 10.1158/1535-7163.MCT-08-0492.

The Gene Ontology Consortium, "Gene Ontology: tool for the unification of biology", *Nat Genet*, 2000 May; 25(1): 25–29. doi: 10.1038/75556, PMCID: PMC3037419, NIHMSID: NIHMS269796, PMID: 10802651

ANEXOS

1- Carga de paquetes a utilizar y obtención de DEGs:

```
1 - #####
2 - #####ANALISIS GENES EXPRESADOS DIFERENCIALMENTE goPROFILES#####
3 - #####
4
5 #Instalamos librerias que vamos a utilizar
6
7 source("http://bioconductor.org/biocLite.R")
8
9 biocLite("limma")
10 biocLite("genefilter")
11 biocLite("gplot")
12 biocLite("ggplot2")
13 biocLite("illuminaMousev2.db")
14 biocLite("AnnotationHub")
15 biocLite("clusterProfiler")
16 biocLite("org.Hs.eg.db")
17 biocLite("DOSE")
18 biocLite("GO.db")
19 biocLite("pathview")
20 biocLite("goProfiles")
21 biocLite("topGO")
22 biocLite("enrichplot")
23 biocLite("igraph")
24 biocLite("devtools")
25 biocLite("Rcpp")
26 biocLite("fgsea")
27 biocLite("ReactomePA")
28 biocLite("pathview")
29 biocLite("KEGGprofile")
30 biocLite("KEGGREST")
31 biocLite("KEGG.db")
32
33 #cargamos las librerias que vamos a utilizar
34
35 library("limma")
36 library("genefilter")
37 library("gplots")
38 library("ggplot2")
39 library("org.Hs.eg.db")
40 library("org.Mm.eg.db")
41 library("illuminaMousev2.db")
42 library("AnnotationHub")
43 library("DOSE")
44 library("enrichplot")
45 library("clusterProfiler")
46 library("GO.db")
47 library("pathview")
48 library("goProfiles")
49 library("topGO")
50 library("igraph")
51 library("devtools")
52 library("Rcpp")
53 library("fgsea")
54 library("ReactomePA")
55 library("pathview")
56 library("KEGGprofile")
57 library("KEGGREST")
58 library("KEGG.db")
59
```

```

60 #abrimos el archivo y le indicamos que el separador es un espacio, sep = " ",
61 #la cabecera son los nombres de los grupos header = TRUE,
62 #el nombre de las columnas es la primera columna, row.names = 1
63 #comprobamos la estructura de nuestra tabla
64
65 raw_data <- read.table("raw.txt", sep = " ", header = TRUE, row.names = 1)
66 str(raw_data)
67 head(raw_data)
68
69 #seleccionamos de nuestra tabla solo los datos de expresión y no los de pval, indicamos la tabla
70 #primer lado fila, segundo columnas. Indicamos que queremos las columnas que corresponden a las muestras wt y a K0,
71
72 ctrl_K0_raw <- raw_data[, c(1, 3, 5, 7, 15, 17)]
73 str(ctrl_K0_raw)
74 head(ctrl_K0_raw)
75
76 #primero leemos el archivo que describe los targets
77
78 targets <- readTargets("targets.txt", row.names = "Name")
79
80 str(targets)
81 head(targets)
82
83 #normalizamos nuestros datos, normalizacion intergrupo
84
85 ctrl_K0_norm_a <- normalizeBetweenArrays(ctrl_K0_raw, method = "quantile")
86 head(ctrl_K0_norm_a)
87 data_frame_norm <- data.frame(ctrl_K0_norm_a)
88
89
90 #pasamos nuestros datos a log2
91
92 ctrl_K0_norm <- log2(ctrl_K0_norm_a)
93 head(ctrl_K0_norm)
94
95 #Filtramos los datos obtenidos por IQR, la varianza que pondremos de corte es 0.5
96 #Rango intercuartil (IQR), medida de dispersión que acompaña a la mediana.
97 #Es la distancia entre los cuartiles superior e inferior
98
99 selected_IQ_data <- varFilter(ctrl_K0_norm, var.func = IQR, var.cutoff = 0.5, filterByQuantile = TRUE)
100 head(selected_IQ_data)
101
102 #realizamos matriz de diseño para indicar cuales son las muestras
103
104 design1 <- cbind (control= c (1,1,1,0,0,0), K0_VACIO= c (0,0,0,1,1,1))
105 rownames(design1) <- targets$Name
106 head(design1)
107
108 #matriz de contraste(comparaciones) para las muestras
109
110 cont.matrix1 <- makeContrasts(controlvsK0_VACIO = K0_VACIO-control, levels = design1)
111 head(cont.matrix1)
112
113 #obtenemos los genes diferencialmente expresados, para obtener un ajuste a un modelo lineal
114
115 fit1 <- lmFit(selected_IQ_data, design1)
116 fit1b <- contrasts.fit(fit1, cont.matrix1)
117 fit1c <- eBayes(fit1b)
118
119
120 #Lista de los genes diferencialmente expresados, metodo de ajuste BH= Benjamini & Hochberg,
121
122 toptableIQR1 <- topTable(fit1c, number = dim(selected_IQ_data)[1], coef = 1, adjust.method = "BH", sort.by = "p")
123
124 #añadimos columna con el nombre de los genes
125
126 gene_names1 <- data.frame(Gene=unlist(mget(row.names(toptableIQR1), envir = illuminaMousev2SYMBOL)))
127 toptableIQR1$gene <- gene_names1$Gene
128 head(toptableIQR1)
129
130 #seleccionamos todos con valor p < 0.05
131
132 significativos1 <- toptableIQR1[toptableIQR1$adj.P.Val < 0.05,]
133 head(significativos1)
134
135 #buscamos los genes que están up-regulados, activados, aquellos que tienen foldchange mayor a 0
136
137 activated1_complete <- significativos1[significativos1$logFC > 1,]
138 head(activated1_complete)
139
140 #hacemos tabla solo con los valores logFC y adj.p.Val
141
142 activated1 <- activated1_complete[, c(1, 5)]
143 head(activated1)

```



```

142
143 #buscamos los genes que estan down-regulados, inhibidos, aquellos que tienen foldchange menor a -1
144
145 inhibited1_complete <- significativos1[significativos1$logFC< -1,]
146 head(inhibited1_complete)
147
148 #hacemos tabla solo con los valores logFC y adj.p.Val
149
150 inhibited1 <- inhibited1_complete[, c(1, 5)]
151 head(inhibited1)
152
153 #hacemos geneList
154
155 #feature 1: numeric vector
156 geneList_act <- activated1_complete[,2]
157 #feature 2: named vector
158 names(geneList_act) <- as.character(activated1_complete[,7])
159 #feature 3: decreasing order
160 geneList_act <- sort(geneList_act, decreasing = TRUE)
161 head(geneList_act)
162
163 #feature 1: numeric vector
164 geneList_inh <- inhibited1_complete[,2]
165 #feature 2: named vector
166 names(geneList_inh) <- as.character(inhibited1_complete[,7])
167 #feature 3: decreasing order
168 geneList_inh <- sort(geneList_inh, decreasing = TRUE)
169 head(geneList_inh)
170
171 #Obtenemos los identificadores de estos genes accediendo
172 #al vector con todos los identificadores row.names con ellos
173 #comprobamos cuántos hay activos y cuántos inactivos con length
174
175 activated.genes1 <- row.names(activated1)
176 length(row.names(activated1))
177 inhibited.genes1 <- row.names(inhibited1)
178 length(row.names(inhibited1))
179
180 #seleccionamos los genes de la lista de activados, eliminamos los NA
181
182 act <- activated1_complete$gene
183 head(act)
184 act_NA <- na.omit(act)
185 head(act_NA)
186
187 #seleccionamos los genes de la lista de inhibidos, eliminamos los NA
188
189 inh <- inhibited1_complete$gene
190 head(inh)
191 inh_NA <- na.omit(inh)
192 head(inh_NA)
193
194 #guardar lista de genes
195
196 write.table(act_NA, file = "act_NA_genes", append = FALSE, quote = TRUE, sep = " ",
197             eol = "\n", na = "NA", dec = ".", row.names = TRUE,
198             col.names = TRUE, qmethod = c("escape", "double"))
199
200 write.table(inh_NA, file = "inh_NA_genes", append = FALSE, quote = TRUE, sep = " ",
201             eol = "\n", na = "NA", dec = ".", row.names = TRUE,
202             col.names = TRUE, qmethod = c("escape", "double"))
203

```


2- Análisis goProfiles:

```
204 ~ #####
205 ~ #####goProfiles#####
206 ~ #####
207
208 #introducimos el keytype GO
209
210 geneID_act_entrez = bitr(act_NA, fromType="SYMBOL", toType=c("ENTREZID"), OrgDb="org.Mm.eg.db")
211 head(geneID_act_entrez)
212
213 #calculamos el basicProfile CC activados
214
215 bP_act_CC <- basicProfile(geneID_act_entrez[,2], idType = "Entrez", onto = "CC", level = 2,orgPackage="org.Mm.eg.db",
216                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
217 bP_act_CC
218 capture.output(bP_act_CC, file="bP_act_CC.doc")
219
220 plotProfiles(bP_act_CC, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
221             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
222             labelWidth=25)
223
224 #calculamos el basicProfile BP activados
225
226 bP_act_BP <- basicProfile(geneID_act_entrez[,2], idType = "Entrez", onto = "BP", level = 2,orgPackage="org.Mm.eg.db",
227                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
228 bP_act_BP
229 capture.output(bP_act_BP, file="bP_act_BP.doc")
230
231 plotProfiles(bP_act_BP, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
232             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
233             labelWidth=25)
234
235 #calculamos el basicProfile MF activados
236
237 bP_act_MF <- basicProfile(geneID_act_entrez[,2], idType = "Entrez", onto = "MF", level = 2,orgPackage="org.Mm.eg.db",
238                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
239 bP_act_MF
240 capture.output(bP_act_MF, file="bP_act_MF.doc")
241
242 plotProfiles(bP_act_MF, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
243             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
244             labelWidth=25)
245
246 #introducimos el keytype GO inhibidos
247 geneID_inh_entrez = bitr(inh_NA, fromType="SYMBOL", toType=c("ENTREZID"), OrgDb="org.Mm.eg.db")
248 head(geneID_inh_entrez)
249
250 #calculamos el basicProfile CC inhibidos
251 bP_inh_CC <- basicProfile(geneID_inh_entrez[,2], idType = "Entrez", onto = "CC", level = 2,orgPackage="org.Mm.eg.db",
252                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
253 bP_inh_CC
254 capture.output(bP_inh_CC, file="bP_inh_CC.doc")
255
256 plotProfiles(bP_inh_CC, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
257             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
258             labelWidth=25)
259
260 #calculamos el basicProfile BP inhibidos
261 bP_inh_BP <- basicProfile(geneID_inh_entrez[,2], idType = "Entrez", onto = "BP", level = 2,orgPackage="org.Mm.eg.db",
262                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
263 bP_inh_BP
264 capture.output(bP_inh_BP, file="bP_inh_BP.doc")
265
266 plotProfiles(bP_inh_BP, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
267             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
268             labelWidth=25)
269
270 #calculamos el basicProfile MF inhibidos
271 bP_inh_MF <- basicProfile(geneID_inh_entrez[,2], idType = "Entrez", onto = "MF", level = 2,orgPackage="org.Mm.eg.db",
272                          ord = TRUE, multilevels = NULL, empty.cats = TRUE, cat.names = TRUE, na.rm = TRUE)
273 bP_inh_MF
274 capture.output(bP_inh_MF, file="bP_inh_MF.doc")
275
276 plotProfiles(bP_inh_MF, aTitle = "Functional Profile", anOnto = NULL, percentage = FALSE,
277             HORIZVERT = TRUE, legendText = NULL, colores = c("white", "red"), multiplePlots = F, multipleWindows = T,
278             labelWidth=25)
279
```

```

280 #comparamos act vs inh
281 #ont= CC
282 compared_act_inh_CC <- compareGeneLists(geneID_act_entrez[,2], geneID_inh_entrez[,2], idType = "Entrez", onto = "CC",
283                                         level = 2, orgPackage="org.Mm.eg.db",
284                                         method = "lcombChisq", ab.approx = "asymptotic", confidence = 0.95,
285                                         compareFunction="compareGOProfiles")
286
287 print(compSummary(compared_act_inh_CC))
288 estadis_act_inh_CC <- compSummary(compared_act_inh_CC)
289 capture.output(estadis_act_inh_CC, file="estadis_act_inh_CC.doc")
290
291 act.inh.CC <- mergeProfilesLists(bP_act_CC, bP_inh_CC, profNames=c("Active", "Inhibited") )
292 act.inh.CC
293 capture.output(act.inh.CC, file="act.inh.CC.doc")
294
295 plotProfiles (act.inh.CC, percentage=T, aTitle="Active vs Inhibited", legend= TRUE, labelWidth=25)
296
297 #ont= BP
298 compared_act_inh_BP <- compareGeneLists(geneID_act_entrez[,2], geneID_inh_entrez[,2], idType = "Entrez", onto = "BP",
299                                         level = 2, orgPackage="org.Mm.eg.db",
300                                         method = "lcombChisq", ab.approx = "asymptotic", confidence = 0.95,
301                                         compareFunction="compareGOProfiles")
302
303 print(compSummary(compared_act_inh_BP))
304 estadis_act_inh_BP <- compSummary(compared_act_inh_BP)
305 capture.output(estadis_act_inh_BP, file="estadis_act_inh_BP.doc")
306
307
308 act.inh.BP <- mergeProfilesLists(bP_act_BP, bP_inh_BP, profNames=c("Active", "Inhibited") )
309 act.inh.BP
310
311 plotProfiles (act.inh.BP, percentage=T, aTitle="Active vs Inhibited", labelWidth=25)
312
313 #ont= MF
314 compared_act_inh_MF <- compareGeneLists(geneID_act_entrez[,2], geneID_inh_entrez[,2], idType = "Entrez", onto = "MF",
315                                         level = 2, orgPackage="org.Mm.eg.db",
316                                         method = "lcombChisq", ab.approx = "asymptotic", confidence = 0.95,
317                                         compareFunction="compareGOProfiles")
318
319 print(compSummary(compared_act_inh_MF))
320 estadis_act_inh_MF <- compSummary(compared_act_inh_MF)
321 capture.output(estadis_act_inh_MF, file="estadis_act_inh_MF.doc")
322
323 act.inh.MF <- mergeProfilesLists(bP_act_MF, bP_inh_MF, profNames=c("Active", "Inhibited") )
324 act.inh.MF
325
326 plotProfiles (act.inh.MF, percentage=T, aTitle="Active vs Inhibited", legend=T, labelWidth=25)
327
328 ~ #####
329 ~ #####

```

3- Análisis clusterProfiler groupGO, enrichGO:

```

218 ~ #####
219 ~ #####clusterProfiler#####
220 ~ #####
221
222 #seleccionamos los genes de la lista de activados, eliminamos los NA
223 act <- activated1_complete$gene
224 head(act)
225
226 act_NA <- na.omit(act)
227 head(act_NA)
228
229 #seleccionamos los genes de la lista de inhibidos, eliminamos los NA
230 inh <- inhibited1_complete$gene
231 head(inh)
232
233 inh_NA <- na.omit(inh)
234 head(inh_NA)
235

```

```

236 #####group GO#####
237
238 #introducimos el keytype GO
239 geneID_act_entrez = bitr(act_NA, fromType="SYMBOL", toType=c("ENTREZID"), OrgDb="org.Mm.eg.db")
240 head(geneID_act_entrez)
241
242 #introducimos el keytype GO inhibidos
243 geneID_inh_entrez = bitr(inh_NA, fromType="SYMBOL", toType=c("ENTREZID"), OrgDb="org.Mm.eg.db")
244 head(geneID_inh_entrez)
245
246 #obtenemos el groupGO de los genes activados, perfil GO, ontologia=CC, celular component
247
248 ggo_act_CC <- groupGO(geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "CC", level = 3, readable = FALSE)
249 head(ggo_act_CC)
250 y <- as.data.frame(ggo_act_CC)
251 capture.output(y , file="y.doc")
252
253 #Hacemos df con datos de Description y Count
254
255 orden_CC_ACT <- as.data.frame(ggo_act_CC$Description)
256 orden_CC_ACT$Count <- ggo_act_CC$Count
257 orden_CC_ACT$porcentaje <- ((orden_CC_ACT$Count)*100/293)
258
259 #hacemos grafica de barras de resultados de groupGO, ont = "CC"
260 barplot(ggo_act_CC, showCategory=30, order=TRUE, main = "goupGO CC ontology active")
261
262 #obtenemos el groupGO de los genes activados, perfil GO, ontologia=BP, biological process
263 ggo_act_BP <- groupGO(geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "BP", level = 3, readable = FALSE)
264 head(ggo_act_BP)
265
266 #Hacemos df con datos de Description y Count
267
268 orden_BP_ACT <- as.data.frame(ggo_act_BP$Description)
269 orden_BP_ACT$Count <- ggo_act_BP$Count
270 orden_BP_ACT$porcentaje <- ((orden_BP_ACT$Count)*100/293)
271
272 #hacemos grafica de barras de resultados de groupGO
273 barplot(ggo_act_BP, showCategory=50, order=TRUE, main = "goupGO BP ontology active")
274
275 #obtenemos el groupGO de los genes activados, perfil GO, ontologia=MF, molecular function
276 ggo_act_MF <- groupGO(geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "MF", level = 3, readable = FALSE)
277 head(ggo_act_MF)
278
279 #Hacemos df con datos de Description y Count
280
281 orden_MF_ACT <- as.data.frame(ggo_act_MF$Description)
282 orden_MF_ACT$Count <- ggo_act_MF$Count
283 orden_MF_ACT$porcentaje <- ((orden_MF_ACT$Count)*100/293)
284
285 #hacemos grafica de barras de resultados de groupGO
286 barplot(ggo_act_MF, showCategory=30, order=TRUE, main = "goupGO MF ontology active")
287
288 #ggo <- GO clasificacion inhibidos, perfil GO, ontologia=CC, celular component
289
290 ggo_inh_CC <- groupGO(geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "CC", level = 3, readable = FALSE)
291 head(ggo_inh_CC)
292
293 barplot(ggo_inh_CC, showCategory=30, order=TRUE, main = "goupGO CC ontology inhibited")
294
295 #Hacemos df con datos de Description y Count
296
297 orden_CC_INH <- as.data.frame(ggo_inh_CC$Description)
298 orden_CC_INH$Count <- ggo_inh_CC$Count
299 orden_CC_INH$porcentaje <- ((orden_CC_INH$Count)*100/396)
300
301 #ggo <- GO clasificacion inhibidos, perfil GO, ontologia=BP, biological process
302
303 ggo_inh_BP <- groupGO(geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "BP", level = 3, readable = FALSE)
304 head(ggo_inh_BP)
305
306 #Hacemos df con datos de Description y Count
307
308 orden_BP_INH <- as.data.frame(ggo_inh_BP$Description)
309 orden_BP_INH$Count <- ggo_inh_BP$Count
310 orden_BP_INH$porcentaje <- ((orden_BP_INH$Count)*100/396)
311
312 barplot(ggo_inh_BP, showCategory=50, order=TRUE, main = "goupGO BP ontology inhibited")
313

```

```

314 #ggo <- GO clasificacion inhibidos, perfil GO, ontologia=MF, molecular function
315
316 ggo_inh_MF <- groupGO((geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db", ont= "MF", level = 3, readable = FALSE)
317 head(ggo_inh_MF)
318
319 #Hacemos df con datos de Description y Count
320
321 orden_MF_INH <- as.data.frame(ggo_inh_MF$Description)
322 orden_MF_INH$Count <- ggo_inh_MF$Count
323 orden_MF_INH$porcentaje <- ((orden_MF_INH$Count)*100/396)
324
325 barplot(ggo_inh_MF, showCategory=30, order=TRUE, main = "goupGO MF ontology inhibited")
326
327
328 #####enrich GO#####
329
330 ego_act_CC <- enrichGO((geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "CC", pAdjustMethod = "BH",
331 pvalueCutoff = 0.05, readable = FALSE)
332 head(ego_act_CC)
333 barplot(ego_act_CC, showCategory=60, main = "enrichGO CC ontology activated")
334
335 ego_act_BP <- enrichGO((geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "BP", pAdjustMethod = "BH",
336 pvalueCutoff = 0.05, readable = FALSE)
337 head(ego_act_BP)
338 barplot(ego_act_BP, showCategory=60, main = "enrichGO BP ontology activated")
339
340 ego_act_MF <- enrichGO((geneID_act_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "MF", pAdjustMethod = "BH",
341 pvalueCutoff = 0.05, readable = FALSE)
342 head(ego_act_MF)
343 barplot(ego_act_MF, showCategory=60, main = "enrichGO MF ontology activated")
344
345 #ego <- ego clasificacion inhibidos
346
347 ego_inh_CC <- enrichGO((geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "CC", pAdjustMethod = "BH",
348 pvalueCutoff = 0.05, readable = FALSE)
349 head(ego_inh_CC)
350 barplot(ego_inh_CC, showCategory=60, main = "enrichGO CC ontology inhibited")
351
352 ego_inh_BP <- enrichGO((geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "BP", pAdjustMethod = "BH",
353 pvalueCutoff = 0.05, readable = FALSE)
354 head(ego_inh_BP)
355 barplot(ego_inh_BP, showCategory=60, main = "enrichGO BP ontology inhibited")
356
357 ego_inh_MF <- enrichGO((geneID_inh_entrez[,2]), OrgDb= "org.Mm.eg.db",keyType = "ENTREZID", ont= "MF", pAdjustMethod = "BH",
358 pvalueCutoff = 0.05, readable = FALSE)
359
360 head(ego_inh_MF)
361 barplot(ego_inh_MF, showCategory=60, main = "enrichGO MF ontology inhibited")
362
363 ###ego-setReadable function to translate geneID to gene symbol#####
364
365 ego1_act_CC <- setReadable(ego_act_CC, OrgDb = org.Mm.eg.db)
366 ego1_act_BP <- setReadable(ego_act_BP, OrgDb = org.Mm.eg.db)
367 ego1_act_MF <- setReadable(ego_act_MF, OrgDb = org.Mm.eg.db)
368 ego1_inh_CC <- setReadable(ego_inh_CC, OrgDb = org.Mm.eg.db)
369 ego1_inh_BP <- setReadable(ego_inh_BP, OrgDb = org.Mm.eg.db)
370 ego1_inh_MF <- setReadable(ego_inh_MF, OrgDb = org.Mm.eg.db)
371
372
373 #####ego graficas#####
374
375 dotplot(ego_act_CC, title = "Active CC")
376 dotplot(ego_act_BP, title = "Active BP")
377 dotplot(ego_act_MF, title = "Active MF")
378 dotplot(ego_inh_CC, title = "Inhibited CC")
379 dotplot(ego_inh_BP, title = "Inhibited BP")
380 dotplot(ego_inh_MF, title = "Inhibited MF")
381
382
383 #####cnetplot ego#####
384 #cnetplot(ego2, foldChange=genelist)
385
386 cnetplot(ego1_act_CC, foldChange = genelist_act)
387 cnetplot(ego1_act_BP, foldChange = genelist_act )
388 cnetplot(ego1_act_MF, foldChange = genelist_act)
389 cnetplot(ego1_inh_CC, foldChange = genelist_inh)
390 cnetplot(ego1_inh_BP, foldChange = genelist_inh)
391 cnetplot(ego1_inh_MF, foldChange = genelist_inh)
392
393 cnetplot(ego1_act_CC, foldChange = genelist_act, circular = TRUE, colorEdge = TRUE)
394 cnetplot(ego1_act_BP, foldChange = genelist_act, circular = TRUE, colorEdge = TRUE)
395 cnetplot(ego1_act_MF, foldChange = genelist_act, circular = TRUE, colorEdge = TRUE)
396 cnetplot(ego1_inh_CC, foldChange = genelist_inh, circular = TRUE, colorEdge = TRUE)
397 cnetplot(ego1_inh_BP, foldChange = genelist_inh, circular = TRUE, colorEdge = TRUE)
398 cnetplot(ego1_inh_MF, foldChange = genelist_inh, circular = TRUE, colorEdge = TRUE)
399
400

```



```

401 ##upsetplot ego#####alternativa a cnetplot
402
403 upsetplot(ego_act_CC)
404 upsetplot(ego_act_BP)
405 upsetplot(ego_act_MF)
406 upsetplot(ego_inh_CC)
407 upsetplot(ego_inh_BP)
408 upsetplot(ego_inh_MF)
409
410 ##emapplot ego#####enrichplot
411
412 emapplot(ego_act_CC)
413 emapplot(ego_act_BP)
414 emapplot(ego_act_MF)
415 emapplot(ego_inh_CC)
416 emapplot(ego_inh_BP)
417 emapplot(ego_inh_MF)
418
419 ###goplot ego#####enrichplot
420
421 goplot(ego_act_CC)
422 goplot(ego_act_BP)
423 goplot(ego_act_MF)
424 goplot(ego_inh_CC)
425 goplot(ego_inh_BP)
426 goplot(ego_inh_MF)
427
428 - #####plotG0graph#####
429
430 plotG0graph(ego_act_CC)
431 plotG0graph(ego_act_BP)
432 plotG0graph(ego_act_MF)
433 plotG0graph(ego_inh_CC)
434 plotG0graph(ego_inh_BP)
435 plotG0graph(ego_inh_MF)
436
437 ##heatplot#####enrichplot
438
439 heatplot(ego1_act_CC, foldChange = geneList_act)
440 heatplot(ego1_act_BP, foldChange = geneList_act)
441 heatplot(ego1_act_MF, foldChange = geneList_act)
442 heatplot(ego1_inh_CC, foldChange = geneList_act)
443 heatplot(ego1_inh_BP, foldChange = geneList_act)
444 heatplot(ego1_inh_MF, foldChange = geneList_act)
445
446 - ##gseG0- Gene Set Enrichment Analysis#####
447
448 ## feature 1: numeric vector
449 geneList_act2 <- activated1_complete[,2]
450 ## feature 2: named vector
451 names(geneList_act2) <- as.character(geneID_act_entrez[,2])
452 ## feature 3: decreasing order
453 geneList_act2 <- sort(geneList_act2, decreasing = TRUE)
454 head(geneList_act2)
455
456 geneList_act2_NA <- na.omit(geneList_act2)
457 head(geneList_act2_NA)
458
459 ## feature 1: numeric vector
460 geneList_inh2 <- inhibited1_complete[,2]
461 ## feature 2: named vector
462 names(geneList_inh2) <- as.character(geneID_inh_entrez[,2])
463 ## feature 3: decreasing order
464 geneList_inh2 <- sort(geneList_inh2, decreasing = TRUE)
465 head(geneList_inh2)
466
467 gsecc_act_CC <- gseG0(geneList_act2, ont = "CC", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
468                      pAdjustMethod = "BH")
469 summary(gsecc_act_CC)
470 head(gsecc_act_CC)
471
472 gsecc_act_BP <- gseG0(geneList_act2, ont = "BP", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
473                      pAdjustMethod = "BH")
474 summary(gsecc_act_BP)
475 head(gsecc_act_BP)
476
477 gsecc_act_MF <- gseG0(geneList_act2, ont = "MF", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
478                      pAdjustMethod = "BH")
479 summary(gsecc_act_MF)
480 head(gsecc_act_MF)
481

```

```

482 gsecc_inh_CC <- gseGO(geneList_inh2, ont = "CC", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
483 pAdjustMethod = "BH")
484 summary(gsecc_inh_CC)
485 head(gsecc_inh_CC)
486
487 gsecc_inh_BP <- gseGO(geneList_inh2, ont = "BP", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
488 pAdjustMethod = "BH")
489 summary(gsecc_inh_BP)
490 head(gsecc_inh_BP)
491
492 gsecc_inh_MF <- gseGO(geneList_inh2, ont = "MF", OrgDb = org.Mm.eg.db, pvalueCutoff= 0.05,
493 pAdjustMethod = "BH")
494 summary(gsecc_inh_MF)
495 head(gsecc_inh_MF)
496
497 ridgeplot(gsecc_act_CC)
498 ridgeplot(gsecc_act_BP)
499 ridgeplot(gsecc_act_MF)
500 ridgeplot(gsecc_inh_CC)
501 ridgeplot(gsecc_inh_BP)
502 ridgeplot(gsecc_inh_MF)
503
504 emapplot(gsecc_act_CC)
505 emapplot(gsecc_act_BP)
506 emapplot(gsecc_act_MF)
507 emapplot(gsecc_inh_CC)
508 emapplot(gsecc_inh_BP)
509 emapplot(gsecc_inh_MF)
510

```

4- Análisis clusterProfiler enrichKEGG:

```

250 #####enrichKEGG#####
251
252 ##convertimos nuestros genes de EntrezID a uniprot y de ahi a KEGG
253
254 act_NA_convert <- bitr((geneID_act_entrez[,2]), fromType="ENTREZID", toType= "ENSEMBL", OrgDb="org.Mm.eg.db")
255 head(act_NA_convert)
256
257 act_NA_sorted <- sort((act_NA_convert[,1]), decreasing = TRUE)
258 head(act_NA_sorted)
259
260 inh_NA_convert <- bitr((geneID_inh_entrez[,2]), fromType="ENTREZID", toType= "ENSEMBL", OrgDb="org.Mm.eg.db")
261 head(inh_NA_convert)
262
263 inh_NA_sorted <- sort((inh_NA_convert[,1]), decreasing = TRUE)
264 head(inh_NA_sorted)
265
266
267 kk_act <- enrichKEGG((act_NA_sorted), organism = "mouse", pvalueCutoff = 0.05)
268 summary(kk_act)
269 kk_inh <- enrichKEGG((inh_NA_sorted), organism = "mouse", pvalueCutoff = 0.05)
270 summary(kk_inh)
271
272 #####geneList2#####
273
274 ## feature 1: numeric vector
275 geneList_act2 <- activated1_complete[,2]
276 ## feature 2: named vector
277 names(geneList_act2) <- as.character(geneID_act_entrez[,2])
278 ## feature 3: decreasing order
279 geneList_act2 <- sort(geneList_act2, decreasing = TRUE)
280 head(geneList_act2)
281
282 geneList_act2_NA <- na.omit(geneList_act2)
283 head(geneList_act2_NA)
284
285 ## feature 1: numeric vector
286 geneList_inh2 <- inhibited1_complete[,2]
287 ## feature 2: named vector
288 names(geneList_inh2) <- as.character(geneID_inh_entrez[,2])
289 ## feature 3: decreasing order
290 geneList_inh2 <- sort(geneList_inh2, decreasing = TRUE)
291 head(geneList_inh2)
292
293 ##setReadable function to translate geneID to gene symbol#####
294
295 kk_act <- setReadable(kk_act, OrgDb = org.Mm.eg.db, keytype = "ENTREZID")
296 kk_inh <- setReadable(kk_inh, OrgDb = org.Mm.eg.db, keytype = "ENTREZID")
297

```

```

298 #####plots#####
299 barplot(kk_act)
300 barplot(kk_inh)
301
302 dotplot(kk_act)
303 dotplot(kk_inh)
304
305 emapplot(kk_act)
306 emapplot(kk_inh)
307
308 cnetplot(kk_act, foldChange = geneList_act2, colorEdge = TRUE)
309 cnetplot(kk_inh, foldChange = geneList_inh2, colorEdge = TRUE)
310
311 cnetplot(kk_act, foldChange = geneList_act2, circular = TRUE, colorEdge = TRUE)
312 cnetplot(kk_inh, foldChange = geneList_inh2, circular = TRUE, colorEdge = TRUE)
313
314 upsetplot(kk_act)
315 upsetplot(kk_inh)
316
317 heatmap(kk_act, foldChange = geneList_act)
318 heatmap(kk_inh, foldChange = geneList_inh)
319
320
321 pathway(gene.data = geneList_act, pathway.id = "mmu05322", species = "mmu",
322         limit = list(gene=max(abs(geneList_act)), cpd=1))
323 pathway(gene.data = geneList_act, pathway.id = "mmu05034", species = "mmu",
324         limit = list(gene=max(abs(geneList_act)), cpd=1))
325
326 pathway(gene.data = geneList_inh, pathway.id = "mmu04210", species = "mmu",
327         limit = list(gene=max(abs(geneList_inh)), cpd=1))
328 pathway(gene.data = geneList_inh, pathway.id = "mmu04070", species = "mmu",
329         limit = list(gene=max(abs(geneList_inh)), cpd=1))
330 pathway(gene.data = geneList_inh, pathway.id = "mmu05222", species = "mmu",
331         limit = list(gene=max(abs(geneList_inh)), cpd=1))
332
333 #####Gene Set Enrichment Analysis#####*ojo pvalue muy alto
334
335 kk_act2 <- gseKEGG(geneList = geneList_act2, organism = "mouse", pvalueCutoff = 0.05)
336 head(kk_act2)
337 kk_inh2 <- gseKEGG(geneList = geneList_inh2, organism = "mouse", pvalueCutoff = 0.05 )
338 head(kk_inh2)
339
340 summary(kk_act2)
341 summary(kk_inh2)
342
343 #####plots#####
344
345 ridgeplot(kk_act2)
346 ridgeplot(kk_inh2)
347
348 gseaplot(kk_act2, geneSetID = "mmu05203")
349 gseaplot(kk_act2, geneSetID = "mmu01100")
350 gseaplot(kk_act2, geneSetID = "mmu05322")
351 gseaplot(kk_act2, geneSetID = "mmu05034")
352 gseaplot(kk_act2, geneSetID = "mmu05200")
353
354 gseaplot(kk_inh2, geneSetID = "mmu05165")
355 gseaplot(kk_inh2, geneSetID = "mmu04510")
356 gseaplot(kk_inh2, geneSetID = "mmu05200")
357 gseaplot(kk_inh2, geneSetID = "mmu01100")
358 gseaplot(kk_inh2, geneSetID = "mmu04151")
359 gseaplot(kk_inh2, geneSetID = "mmu04010")
360 gseaplot(kk_inh2, geneSetID = "mmu04210")
361
362 emapplot(kk_act2)
363 emapplot(kk_inh2)
364
365 dotplot(kk_act2)
366 dotplot(kk_inh2)
367
368 cnetplot(kk_act2, foldChange = geneList_act2, circular = TRUE, colorEdge = TRUE)
369 cnetplot(kk_inh2, foldChange = geneList_inh2, circular = TRUE, colorEdge = TRUE)
370
371
372 #####KEGG Module over-representation test#####no sale nada solo mkk_act
373
374 mkk_act <- enrichMKEGG((act_NA_sorted), organism = "mmu", pvalueCutoff = 0.05)
375 summary(mkk_act)
376 mkk_inh <- enrichMKEGG((inh_NA_sorted), organism = "mmu", pvalueCutoff = 0.05)
377 summary(mkk_inh)
378
379 mkk2_act <- gseMKEGG(geneList = geneList_act2, organism = "mmu", pvalueCutoff = 0.05)
380 summary(mkk2_act)
381 mkk2_inh <- gseMKEGG(geneList = geneList_inh2, organism = "mmu", pvalueCutoff = 0.05)
382 summary(mkk2_inh)
383
384 #####

```

5- Análisis reactomePA:

```
221 #####
222 #####ReactomePA#####
223 #####
224
225 act_NA_convert <- bitr(act_NA, fromType="SYMBOL", toType= "ENTREZID", OrgDb="org.Mm.eg.db")
226 head(act_NA_convert)
227 geneList_actK <- act_NA_convert[,2]
228 head(geneList_actK)
229
230 enr_P_act <- enrichPathway (geneList_actK, organism = "mouse", pvalueCutoff=0.05, pAdjustMethod = "BH", readable = FALSE)
231 head(enr_P_act)
232 summary(enr_P_act)
233
234 inh_NA_convert <- bitr(inh_NA, fromType="SYMBOL", toType= "ENTREZID", OrgDb="org.Mm.eg.db")
235 head(inh_NA_convert)
236 geneList_inhK <- inh_NA_convert[,2]
237 head(geneList_inhK)
238
239 enr_P_inh <- enrichPathway (geneList_inhK, organism = "mouse", pvalueCutoff=0.05, pAdjustMethod = "BH", readable = FALSE)
240 head(enr_P_inh)
241 summary(enr_P_inh)
242
243 ###barplot#####
244
245 barplot(enr_P_act,drop=TRUE, showCategory=60, main = "enrichPathway active")
246 barplot(enr_P_inh,drop=TRUE, showCategory=15)
247
248 ###dotplot#####
249
250 dotplot(enr_P_act, showCategory=15)
251 dotplot(enr_P_inh, showCategory=15)
252
253 ###cnetplot#####
254
255 enr1_P_act <- setReadable(enr_P_act, OrgDb = org.Mm.eg.db)
256 head(enr1_P_act)
257 cnetplot(enr1_P_act, categorySize="pvalue", foldChange=geneList_act)
258
259 enr1_P_inh <- setReadable(enr_P_inh, OrgDb = org.Mm.eg.db)
260 head(enr1_P_inh)
261 cnetplot(enr1_P_inh, categorySize="pvalue", foldChange=geneList_inh)
262
263 cnetplot(enr1_P_act, foldChange = geneList_act, circular = TRUE, colorEdge = TRUE)
264 cnetplot(enr1_P_inh, foldChange = geneList_act, circular = TRUE, colorEdge = TRUE)
265
266 ###emapplot#####
267
268 emapplot (enr1_P_act)
269 emapplot (enr1_P_inh)
270
271 emapplot(enr_P_act)
272
273 #####Gene Set Enrichment Analysis#####
274
275 ## feature 1: numeric vector
276 geneList_act2 <- activated1_complete[,2]
277 ## feature 2: named vector
278 names(geneList_act2) <- as.character(geneID_act_entrez[,2])
279 ## feature 3: decreasing order
280 geneList_act2 <- sort(geneList_act2, decreasing = TRUE)
281 head(geneList_act2)
282
283 geneList_act2_NA <- na.omit(geneList_act2)
284 head(geneList_act2_NA)
285
286 gse_act <- gsePathway( geneList_act2, nPerm=10000,
287                       pvalueCutoff=0.2,
288                       pAdjustMethod="BH", verbose=FALSE, organism = "mouse")
289 gse_act_df <- as.data.frame(gse_act)
290 head(gse_act_df)
291
292 emapplot(gse_act, color="pvalue")
293
294 gseaplot(gse_act, geneSetID = "R-MMU-212436")
295 gseaplot(gse_act, geneSetID = "R-MMU-556833")
296
297
298 viewPathway("E2F mediated regulation of DNA replication", readable=TRUE, foldChange=geneList)
299
300
301 #####
302 #####
```